Einführung in die Stochastik für Informatiker
Professor Kamps
SS 2004

Dominik Lenhard (Dominik@DLenhard.de)

12. Juli 2004

Ein Dankeschön geht an dieser Stelle
an Peter Schroeder, der die Skizzen angefertigt hat
und an Christian Köhler,
der mich zum Teil mit Mitschriften versorgt hat.

Inhaltsverzeichnis

0 Einführung

Ziele und Aufgaben der Stochastik:

- Modelle für reale, zufallsabhängige Vorgänge
  - Aussagen im Modell
  - Entscheidungshilfe durch „Rückübersetzung“ in die Realität
- Datenanalyse

Spezielle Beispiele:

- Netzwerkanalyse
- Warteschlangen
- Analyse von Algorithmen
- Simulation von Systemen

Deterministische Vorgänge (Ursache $\leftrightarrow$ Wirkung) sind eine Seite.
Zufallsabhängige Vorgänge die Andere:

- prinzipiell ein Ergebnis nicht vorhersehbar, weil „Zufall eingreift“
  - Ergebnisse von Glücksspielen
  - Lebensdauer eines Systems
  - falsch übertragene Bits
- Vorhersage prinzipiell möglich, aber zu komplex
- Zahl der Einflussgrößen zu hoch
- Probleme bei Quantifizierung der Eingabegrößen

Ziel:
Sicherheit über die Unsicherheit gewinnen im folgenden Sinn: Einzelversuch nicht vorhersagbar, aber Aussage bei häufiger Versuchswiederholung möglich. $\hookrightarrow$ Datenerhebung, Simulation

Beispiel:
Würfelwurf als Experiment:
$\hookrightarrow$ mögliche Ergebnisse kodiert durch $1, ..., 6$.
Stichprobenraum: $\Omega = \{1, ..., 6\}$.
Ergebnis: $\omega \in \Omega$ (Elementarereignis)
„Klar“: gleiche Chance für jede Zahl: $P(\{\omega\}) = \frac{1}{6}\ \forall \omega \in \Omega$.

Andere Fragestellung:
„Es fällt eine gerade Zahl.“
$\hookrightarrow$ beschreibbar durch Teilmenge $A := \{2, 4, 6\} \subseteq \Omega$
„Klar“: $P(A) = \frac{1}{2}$
$\hookrightarrow$Modell: Ω, $\mathcal{P}(\Omega) = \{A | A \subseteq \Omega\}$, $P : \mathcal{P}(\Omega) \to [0, 1]$ mit $P(A) = \frac{|A|}{6}$ $\forall A \subset \Omega$.
Zu überprüfen:
Übereinstimung von Modell mit Realität durch Experiment $\to$ Statistik.

1 Diskrete Wahrscheinlichkeitsräume

1.1 Diskrete Wahrscheinlichkeitsräume und deren Erweiterung

(1.1) **Definition**: Sei Ω eine höchstabzählbare Menge. $\mathfrak{A} := \mathcal{P}(\Omega) := \{A | A \subseteq \Omega\}$ und $p : \Omega \to [0,1]$ eine Abbildung mit $\sum_{\omega \in \Omega} p(\omega) = 1$. Die durch $P(A) = \sum_{\omega \in A} p(\omega)$, $A \subseteq \Omega$, definierte Abbildung heißt (diskrete) Wahrscheinlichkeitsverteilung (WV) über Ω. Die Funktion p heißt Zähldichte. Das Tripel $(\Omega, \mathfrak{A}, P)$ heißt diskreter Wahrscheinlichkeitsraum. $(\Omega, \mathfrak{A})$ heißt diskreter messbarer Raum.

(1.2) **Bezeichnung**: Ω heißt Grundmenge, Ergebnisraum, Stichprobenraum. $A \subseteq \Omega$ heißt Ereignis und speziell Elementarereignis, falls $|A| = 1$. Kurzschreibweise: $P(\{\omega\}) = P(\omega) = p(\omega), \omega \in \Omega$.

(1.3) **Lemma**:

a) Gegeben sei die Situation aus (1.1): Dann gilt:
   i) $0 \leq P(A) \leq 1 \; \forall A \in \mathcal{P}(\Omega)$
   ii) $P(\Omega) = 1$
   iii) P ist σ-additiv, das heißt, für paarweise disjunkte $A_1, A_2, ... \in \mathcal{P}(\Omega)$ (das heißt, $A_i \cap A_j = \emptyset$ für $i \neq j$) gilt: $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. Insbesondere ist $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{u} P(A_i) \; \forall u \in \mathbb{N}$ und $P(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1$.

b) Sei $P : p(\omega) \to \mathbb{R}^1$ eine Abbildung mit den oben genannten drei Bedingungen. $\Rightarrow$ Es gibt genau eine Funktion $p : \Omega \to [0,1]$ mit $P(A) = \sum_{\omega \in A} p(\omega) \; \forall A \subset \Omega$.

Beweis:

a) iii) Dritte Bedingung:

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{\omega \in \bigcup_{i=1}^{\infty} A_i} P(\omega) = \sum_{i=1}^{\infty} \sum_{\omega \in A_i} P(\omega) = \sum_{i=1}^{\infty} P(A_i).$$

(1.4) **Beispiel**: Der Laplaceraum.
Sei $\Omega = \{1, ..., n\}$, $p(\omega) = \frac{1}{n} \; \forall \omega \in \Omega$. P heißt dann diskrete Gleichverteilung oder Laplaceverteilung.
Es ist $P(A) = \frac{|A|}{|\Omega|} = \frac{1}{n} \cdot |A| \; \forall A \subseteq \Omega$.
$\hookrightarrow$„Anzahl der günstigen Fälle dividiert durch die Anzahl aller Möglichkeiten.“

(1.5) **Beispiel**: Ein Problem des Chavalier de Niere.
Was ist bei 3 Würfelwürfen wahrscheinlicher? Summe 11 oder Summe 12?
Von Interesse:

– Ereignis A: Summe der Augen = 11,
– Ereignis B: Summe der Augen = 12.

Jeweils 6 Fälle.
Modell: $\Omega = \{\omega = \{\omega_1, \omega_2, \omega_3\};\, \omega_i \in \{1, ..., 6\}\}$.
$A = \{\omega \in \Omega;\, \omega_1 + \omega_2 + \omega_3 = 11\}$
$B = \{\omega \in \Omega;\, \omega_1 + \omega_2 + \omega_3 = 12\}$
$p(\omega) = \frac{1}{\Omega} = \frac{1}{6^3}\ \forall \omega \in \Omega$ (Symmetrie)
Also: Ω ist eine Menge von Tripeln, das heißt, Würfel sind unterscheidbar, das heißt, $(6,4,1) \neq (6,1,4) \neq (1,4,6)$.
Abzählen liefert: $|A| = 27$, $|B| = 25$.
Also: $P(A) = \sum_{\omega \in A} p(\omega) = \frac{27}{216} > \frac{25}{216} = P(B)$.

Auf Laplaceräumen:
Viele Argumente der Wahrscheinlichkeitsrechnung reduzieren sich auf Kombinatorik und „geschicktes“Abzählen.
$\hookrightarrow$ Wahl der „richtigen“ Modelle
$\hookrightarrow$ Wahl der „richtigen“ Zählweise
i.a. Inhalt von Kapitel 2.

Zunächst noch Erweiterung und Ausblick:
Ausgangspunkt: σ-Additivität in (1.3).
Zur Modellierung gewünscht:
$\Omega = [0,1]$ (Zinssatz)
$\Omega = \mathbb{R}$ (Abweichung von Sollwert)
$\Omega = \mathbb{R}^+$ (Lebensdauer eines Systems)
$\hookrightarrow$ Probleme, falls Ω überabzählbar.
Es gibt keine „Gleichverteilung“ über die Potenzmenge von $[0,1]$, das heißt, es gibt keine WV mit der Eigenschaft $P(A+h) = P(A)\ \forall A \in \mathcal{P}(\Omega)$, $A+h \in \mathcal{P}(\Omega)$, wobei $A + h := \{a+h, a \in A\}, h \in \mathbb{R}$.
Also (Konsequenz aus der Mathematik):
Beschränkung auf geeignetes Mengensystem, das echt enthalten ist in der Potenzmenge $\mathcal{P}(\Omega)$ (falls σ-Additivität weiter gefordert ist).
Geeigenete Struktur: σ-Algebra.

(1.6) **Definition**: Seien $\Omega \neq \emptyset$ und $\mathfrak{A} \subseteq \mathcal{P}(\Omega)$ (System von Teilmengen von Ω). $\mathfrak{A}$ heißt σ-Algebra (von Ereignissen) über Ω, falls gilt:

i) $\Omega \in \mathfrak{A}$,

ii) $A \in \mathfrak{A} \to A^C \in \mathfrak{A}\ \forall A \in \mathfrak{A}$,

iii) $(A_n)_{n \in \mathbb{N}} \Rightarrow \bigcup_{i=1}^{\infty} A_n \in \mathfrak{A}$.

Das heißt: Eine σ-Algebra ist abgeschlossen gegenüber der Bildung von Komplementen und abzählbaren Vereinigungen.
Bezeichnung: Als Elementarereignis (vgl. (1.3)) bezeichnet man in diesem Zusammenhang eine Menge aus $\mathfrak{A}$, die keine echte Vereinigung anderer Ereignisse ist.

(1.7) **Lemma**:

a) Sei $\mathfrak{A}$ eine σ-Algebra. Dann gilt:
   i) $\emptyset \in \mathfrak{A}$, $(\emptyset = \Omega^C \in \mathfrak{A})$
   ii) $A, B \in \mathfrak{A} \Rightarrow A \cap B \in \mathfrak{A}$
   $(A \cap B) = (A^C \cup B^C)^C$
   iii) $(A_n)_{n\in \mathbb{N}} \subset \mathfrak{A} \Rightarrow \bigcap_{i=1}^{\infty} A_n \in \mathfrak{A}$.

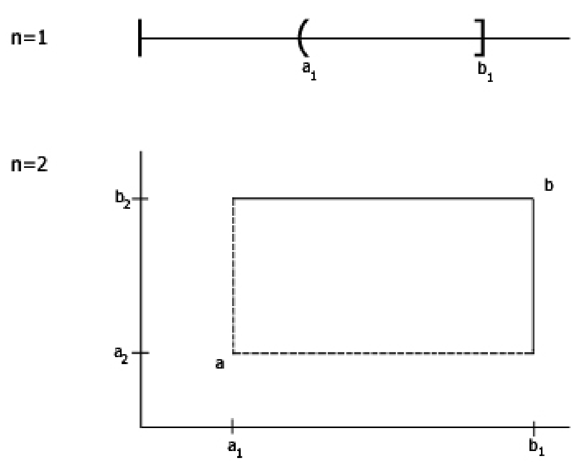
b) Seien $B \subset \Omega$ und $\mathfrak{A}$ eine σ-Algebra über Ω. Dann ist $B \cap \mathfrak{A} := \{B \cap A, A \in \mathfrak{A}\}$ eine σ-Algebra über B (Spur-σ-Algebra).

c) Sei $\Omega \neq \emptyset$.
   $\mathfrak{A} = \{A \subseteq \Omega,\ A$ ist höchstens abzählbar oder A^C ist höchstens abzählbar$\}$
   ist eine σ-Algebra über Ω.

Bemerkung / Bezeichnung:
$\mathcal{P}(\Omega)$ ist σ-Algebra über Ω;
$\mathcal{P}(\Omega)$ ist die feinste, $\mathfrak{A} = \{\emptyset, \Omega\}$ ist die gröbste σ-Algebra über Ω.

Man kann zeigen: Es gibt stets eine σ-Algebra, die ein vorgegebenes System $\mathcal{F}$ von „einfachen“ Mengen enthält.
Genauer:

(1.8) **Satz**: Seien $\Omega \neq \emptyset, \mathcal{F} \subset \mathcal{P}(\Omega)$.
Die kleinste σ-Algebra, die $\mathcal{F}$ enthält, ist gegeben durch $\mathfrak{A}(\mathcal{F}) := \{A \in \mathcal{P}(\Omega)$; für jede σ-Algebra $\mathfrak{A}$ mit $\mathcal{F} \subset \mathfrak{A}$ gilt: $A \in \mathfrak{A}\}$.
$\mathfrak{A}(\mathcal{F})$ heißt die von $\mathcal{F}$ erzeugte σ-Algebra.
Oft zur Modellierung benötigt: $\Omega = \mathbb{R}^n$.
Dann: Wähle $\mathcal{F}$ als Menge aller nach links offenen Intervalle $(a, b] = \{x = (x_1, ..., x_n) \in \mathbb{R}^n; a_i < x_i \leq b_i, 1 \leq i \leq n\}$ für $a = (a_1, ..., a_n)$, $b = (b_1, ..., b_n) \in \mathbb{R}^n$.



Zugehörige σ-Algebra: $\mathfrak{A}(\mathcal{F}) =: \mathfrak{B}^n$.
Borel'sche σ-Algebra.
Bemerkung: Alle offenen und abgeschlossenen Mengen liegen in $\mathfrak{B}^n$, aber es gilt: $\mathfrak{B}^n \neq \mathcal{P}(\mathbb{R}^n))$.

Nun Definition eines allgemeinen Wahrscheinlichkeitsraumes ($\to$ Kolmogorow).

(1.9) **Definition**: Sei $\mathfrak{A}$ eine σ-Algebra über $\Omega \neq \emptyset$.
Eine Abbildung $P : \mathfrak{A} \to [0, 1]$ mit

(i) $P(A) \geq 0 \; \forall A \in \mathfrak{A}$,

(ii) $P(\Omega) = 1$,

(iii) $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ für alle paarweise disjunkten $A_1, A_2, ... \in \mathfrak{A}$ (σ-Additivität),

heißt Wahrscheinlichkeitsverteilung oder Wahrscheinlichkeitsmaß auf $\mathfrak{A}$. $(\Omega, \mathfrak{A}, P)$ heißt Wahrscheinlichkeitsraum, $(\Omega, \mathfrak{A})$ heißt messbarer Raum.

Bemerkung: Vergleiche diskreter Wahrscheinlichkeitsraum: Wahrscheinlichkeitsverteilung ist festgelegt durch die Wahrscheinlichkeiten $P(\omega)$ der Elementarereignisse. Zugang über allgemeine σ-Algebren ist im folgenden Sinne nicht schwieriger: Ist $P(A)$ festgelegt für alle $A \in \mathcal{F}$ mit (i) bis (iii) $\Rightarrow$ (Maßtheorie) $P(B)$ eindeutig bestimmt für alle $B \in \mathfrak{A}(\mathcal{F})$.

Zur besseren Vorstellung und Vereinfachung im Folgenden zunächst: Ω höchstens abzählbar und $\mathfrak{A} = \mathcal{P}(\Omega)$.
Aber: Die im Kapitel 3 hergeleiteten Regeln mit Wahrscheinlichkeiten gelten allgemein.
$\hookrightarrow$ Genaueres später!

Einfachste diskrete Wahrscheinlichkeitsverteilung:

(1.10) Sei $\Omega \neq \emptyset$ abzählbar, $\omega \in \Omega$ fest.
Die durch $\varepsilon_\omega(A) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \notin A \end{cases}, A \in \mathcal{P}(\Omega)$
festgelegte Wahrscheinlichkeitsverteilung $\varepsilon_\omega : \mathcal{P}(\Omega) \to \mathbb{R}$ heißt Dirac-Verteilung oder Einpunktverteilung im Punkt ω.

Bemerkung: ε_ω ist Wahrscheinlichkeitsverteilung, denn:

(i) $P(A) \geq 0$. $\surd$

(ii) $P(\Omega) = 1$, da $\omega \in \Omega$. $\surd$

(iii) $(A_n)_n \subset \mathfrak{A}$ paarweise disjunkt
$\Rightarrow$ es gibt höchstens ein $i \in \mathbb{N} : \omega \in A_i$
$\Rightarrow \varepsilon_\omega(\bigcup_{i=1}^{\infty} A_i) = \begin{cases} 1 & \text{falls } \omega \in \bigcup_{n=1}^{\infty} A_n \\ 0 & \text{sonst} \end{cases}$
$= \begin{cases} 1 & \text{falls es Index } i \text{ gibt: } \omega \in A_i \\ 0 & \text{sonst} \end{cases}$
$= \sum_{n=1}^{\infty} \varepsilon_\omega(A_n)$. $\surd$

(1.11) **Definition**: Sei $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitsraum.
$T := \text{supp}(P) := \{\omega \in \Omega; P(\omega) > 0\}$
heißt Träger von P.

(1.12) **Lemma**: Sei $(\Omega, \mathfrak{A}, P)$ ein diskreter WR und T der Träger von P.
Dann gilt:
$P(A) = \sum_{\omega \in T} P(\omega)\varepsilon_\omega(A)$, $A \in \mathfrak{A}$, das heißt, P ist darstellbar als gewichtete Summe von Einpunktverteilungen.
Beweis:

$$\sum_{\omega \in T} P(\omega)\varepsilon_\omega(A) = \sum_{\omega \in \Omega} P(\omega)\varepsilon_\omega(A) =$$

$$\sum_{\omega \in A} P(\omega)\underbrace{\varepsilon_\omega(A)}_{1} + \sum_{\omega \in A^C} P(\omega)\underbrace{\varepsilon_\omega(A)}_{0} = \sum_{\omega \in A} P(\omega) = P(A).$$

1.2 Grundformeln der Kombinatorik

Kombinatorik: „Kunst des Zählens“
hier: Bestimmung der Mächtigkeit von Mengen.
Grundvoraussetzung: Laplaceraum.

(2.1) **Beispiel**:
Speisekarte: 4 Vorspeisen, 3 Hauptgerichte, 3 Nachspeisen. Anzahl unterschiedlicher Menus? klar: $4 \cdot 3 \cdot 3 = 36$

(2.2) **Beispiel**:
4 Kneipen, Besuchsplan für die nächsten 3 Tage (jeden Tag ein Besuch)

(i) mehrfacher Besuch möglich: $4^3 = 64$ Pläne

(ii) jede Kneipe höchstens einmal: $4 \cdot 3 \cdot 2 = 24$ Pläne

Nun systematisch:

(2.3) **Bemerkung**:
Urnenmodell als Hilfsmittel zum Laplace-Experiment: Urne mit n nummerierten Kugeln (Nummern $1, ..., n$)
$\rightarrow$ sukzessives, zufälliges Ziehen von k Kugeln.
Daher:

- Ziehen
  - mit Zurücklegen (mit Wiederholung)
  - ohne Zurücklegen (ohne Wiederholung)
- Stichprobe
  - mit (zeitlicher) Reihenfolge ($\rightarrow$ Tupel)
  - ohne Reihenfolge ($\hookrightarrow$ lexikografische Ordnung)

(2.4) **Ziehen mit Zurücklegen in Reihenfolge**
Realisierung ist Tupel $\omega = (\omega_1, ..., \omega_k)$, wobei ω_j: Nummer der j-ten gezogenen Kugel.
Aus Symmetriegründen: jedes ω ist gleich wahrscheinlich.
Stichprobenraum: $\Omega_1 = \{\omega = (\omega_1, ..., \omega_k); \omega_i \in \{1, ..., n\}, i \in \{1, ..., k\}\} = \{1, ..., n\}^k$ mit $|\Omega_1| = n^k$ (vergleiche Bsp. (2.2) (i)).

(2.5) **Ziehen ohne Zurücklegen in Reihenfolge**
Realisierung ist k-Tupel mit verschiedenen Einträgen $\Omega_2 = \{\omega \in \{1, ..., n\}^k; \omega_i \neq \omega_j$ für $1 \leq i \neq j \leq k\}$
$|\Omega_2| = n \cdot (n-1) \cdot ... \cdot (n-k+1) = \frac{n!}{(n-k)!} =: (n)_k$
Speziell: $n = k \Rightarrow |\Omega_2| = n!$ und Ω_2 ist Menge aller Permutationen von $1, ..., n$ (vergleiche Bsp. (2.2) (ii)).

(2.6) **Ziehen ohne Zurücklegen ohne Reihenfolge**
Ergebnis einer Ziehung: Welche Kugeln wurden ohne Beachtung der Reihenfolge gezogen?
Daher lexikografische Ordnung:
$\Omega_3 = \{\omega \in \{1, ..., n\}^k; \omega_1 < \omega_2 < ... < \omega_k\}$
oder alternativ
$\Omega_3' = \{A \subset \{1, ..., n\}; |A| = k\}$
Bijektion zwischen Ω_3 und Ω_3': $\omega \mapsto \{\omega_1, ..., \omega_k\})$
$|\Omega_3| = \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} (= |\Omega_3'|)$, denn:
Betrachte Abbildng $f : \Omega_2 \to \Omega_3$ mit $(\omega_1, ..., \omega_k) \mapsto \{\omega_1, ..., \omega_k\}$.
Jedes Urbild $f^{-1}(y) := \{\omega; f(\omega) = y\}$ von $y \in \Omega_3$ hat genau $k!$ Elemente ((2.5) mit $n = k$)
$\Rightarrow |\Omega_3| = \frac{|\Omega_2|}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$.
Merkregel: Die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge ist $\binom{n}{k}$.

(2.7) **Ziehen mit Zurücklegen ohne Reihenfolge**
Realisierung ist k-Tupel mit aufsteigend geordneten Koordinaten.
$|\Omega_4| = \{\omega \in \{1, ..., n\}^k, \omega_1 \leq \omega_2 \leq ... \leq \omega_k\}$ mit $|\Omega_4| = \binom{n+k-1}{k}$.
Denn betrachte Abbildung $(\omega_1, ..., \omega_k) \mapsto (\omega_1, \omega_2 + 1, ..., \omega_k + k - 1)$.
Dies ist bijektive Abbildung von Ω_4 in einen neuen Stichprobenraum.
$\widetilde{\Omega_4}$ eines Modells, wo aus $n+k-1$ nummerischen Kugeln k Kugeln ohne Zurücklegen und ohne Reihenfolge gezogen werden, genauer:
$\widetilde{\Omega_4} = \{(\omega_1', ..., \omega_k') \in B^k; \omega_1' < \omega_2' < ... < \omega_k'\}$ mit $B = \{1, ..., n + k - 1\}$ und $\underbrace{\omega_i'}_{\text{alle verschieden}} = \omega_i + i - 1$
$\underset{(2.6)}{\Rightarrow} |\Omega_4| = |\widetilde{\Omega_4}| = \binom{n+k-1}{k}$.

(2.8) **andere Bezeichnungen**
(2.4): (n, k)-Permutation aus Ω mit Wiederholung.
(2.5): (n, k)-Permutation aus Ω ohne Wiederholung.
(2.6): (n, k)-Kombination aus Ω ohne Wiederholung.
(2.7): (n, k)-Kombination aus Ω mit Wiederholung.

(2.9) **Satz**: Die Urnenmodelle (2.4), (2.5) und (2.6) liefern Laplace-Räume, (2.7) bildet <u>keinen</u> Laplace-Raum.

(2.10) zu (2.7): (Gegen-) Beispiel:
Argumentation über Ziehen mit Zurücklegen in Reihenfolge.
Seien $n = 4$, $k = 4$ (mit Zurücklegen ohne Reihenfolge).
$P((1, 1, 1, 1)) = \frac{1}{4^4}$, aber $P((1, 2, 3, 4)) = \frac{4!}{4^4}$.
Denn: Im Modell mit Reihenfolge ist $(\Omega_1, \mathcal{P}(\Omega_1), P)$ ein Laplace-Raum.

Übersicht	**k-mal ziehen aus n Kugeln**	**mit Zurücklegen**	**ohne Zurücklegen**	
Permutation	mit Reihenfolge	(2.4) Laplace-Raum, $\lvert\Omega_1\rvert = n^k$	(2.5) Laplace-Raum, $\lvert\Omega_2\rvert = (n)_k$	unterscheidbare Murmeln
Kombination	ohne Reihenfolge	(2.7) <u>kein</u> Laplace-Raum, $\lvert\Omega_4\rvert = \binom{n+k-1}{k}$	(2.6) Laplace-Raum, $\lvert\Omega_3\rvert = \binom{n}{k}$	ununterscheidbare Murmeln
		mit Mehrfach-belegung	ohne Mehr-fachbelegung	k Murmeln verteilt auf n Urnen

Dazu:

(2.11) **Interpretation über das Verteilen von Murmeln**:
Anzahl der Möglichkeiten, k Murmeln auf n Plätze zu verteilen.
Murmeln unterscheidbar: i-te Murmel in Zelle ω_i, $1 \leq i \leq k$, $\omega_i \in \{1, ..., n\}$
Nummer der Ziehung entspricht Nummer der Murmel.
Nummer der Kugel entspricht Nummer der Zelle.
$\hookrightarrow$ Fall Ω_1 bei Mehrfachbelegung u.s.w.

(2.12) **Beispiel**:

(i) Geburtstagsproblem:
Gesucht: Wahrscheinlichkeit, mit der von k zufällig ausgewählten Personen mindestens 2 an demselben Tag eines Jahres Geburtstag haben $=: p_k$.
Modell: Jahr mit 365 Tagen, Geburten gleichwahrscheinlich an allen Tagen des Jahres.
$\hookrightarrow \Omega_1$ mit $n = 365$ und n^k Elementen (Ziehen von k Daten mit Zurücklegen).
Ereignis E_k: kein Geburtstag ist doppelt.
(Bekannt: $P(A)+P(A^C) = 1$), das heißt, $P(E_k) = \frac{(n)_k}{n^k} = \frac{(365)_k}{365^k} = q_k = (1-p_k)$
Speziell: $p_{10} \approx 0,12$, $p_{23} \approx 0,51$, $p_{50} \approx 0,97$.

(ii) Hashing (Thematische Datenstrukturen):
Problem:

* Teilmengen eines Ganzen mit Hilfe geeigneter (Hash-) Tafeln abspeichern.
* Zugriff auf derartige Teilmengen.

Vorgehen: Zufälliges Ablegen von k Dateien in einem Feld der Länge n ($\geq k$).
möglich: Mehrfachbelegung $\rightarrow$ Kollision.
Sei $A_{n,k}$ die Bezeichnung für: Eine Kollision findet statt.
$\Rightarrow P(A^C_{n,k}) = \frac{(n)_k}{n^k} = \prod_{i=1}^{k-1}(1 - \frac{i}{n})$.

(2.13) **Beispiel Lotto**:
$k = 6$ Kugeln aus $n = 49$ ohne Zurücklegen:

$$\left.\begin{array}{lll} \text{mit Reihenfolge} & \rightarrow & \Omega_2 \\ \text{ohne Reihenfolge} & \rightarrow & \Omega_3 \end{array}\right\} \text{„Freiheit“ der Modellwahl.}$$

Aber: Fragestellung „3. gezogene Kugel ist die 49“ lässt sich nur in Ω_2 behaupten.
Sei $p_k := P(k \text{ Richtige})$.
$p_6 = \frac{1}{\binom{49}{6}} = \frac{1}{13983816}$
Zu p_4 in Ω_3; „günstiges“ Ereignis A_4.
Sei $\omega' = \{\omega'_1, ..., \omega'_6\}$ die Menge der geratenen Zahlen.
$A_4 = \{\omega \in \Omega_3; |\{\omega_1, ..., \omega_k\} \cap \{\omega'_1, ..., \omega'_6\}| = 4\}$
Dazu: Lege 4 Kugeln von ω' fest (die „richtigen Kugeln“, $\binom{6}{4}$ als Anzahl der Möglichkeiten) und dann 2 Kugeln von $\{1, ..., 49\}\backslash\omega'$ (dies sind $\binom{43}{2}$).
$\hookrightarrow$ alle Kombinationen sind möglich $\Rightarrow \binom{6}{4} \cdot \binom{43}{2}$ ist die Anzahl der günstigen Möglichkeiten.
$\Rightarrow p_4 = \frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} = ...$

(2.14) **Beispiel**:
(2.13) ist Spezialfall der hypergeometrischen Verteilung.
Urne mit S schwarzen und W weißen Kugeln, $n = S + W$; ziehe $k (\leq n)$ Kugeln ohne Zurücklegen. Dann ist die Wahrscheinlichkeit, dass die Stichprobe genau s schwarze und $k - s$ weiße Kugeln enthält:

$$h(s; k, n, S) = \frac{\binom{S}{s}\binom{W}{w}}{\binom{S+W}{k}} = \frac{\binom{S}{s}\binom{n-S}{k-s}}{\binom{n}{k}}, s \leq k.$$

(Argumentation wie in (2.13))
Die durch h bestimmte Zähldichte $(h(0) + h(1) + ... + h(k) = 1)$ definiert die hypergeometrische Verteilung.

Anwendung: Qualitätskontrolle
Warenstichprobe bei Gut-Schlecht-Prüfung
Lieferung von n Teilen (intakt oder defekt) mit S defekten Teilen und $n - S$ intakten Teilen.
Kontrolle: Stichprobe vom Umfang k $(\leq n)$ ohne Zurücklegen.
Damit: Wahrscheinlichkeit für s defekte Teile in der Stichprobe: $h(s; k, n, S)$.

Anwendung: Fische zählen.
Frage: Anzahl der Fische in einem Teich.
Fange S Fische.

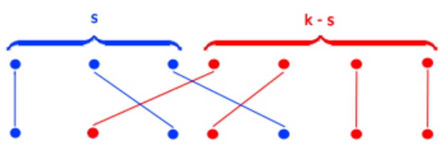
- $\rightarrow$ markiere diese und setze wieder ein.
- $\rightarrow$ warten (d.h. „mischen“).
- $\rightarrow$ fange k Fische, darunter seien s bereits markierte.

($\rightarrow$ in Biologie: capture-recapture-Verfahren)
Intuitiv: Verhältnis $\frac{S}{n} \approx \frac{s}{k}$, d.h., „Schätzer“ für n: $k \cdot \frac{S}{s}$.
Im Rahmen der Mathematischen Statistik:

$h(s; k, n, S)$ ist die Wahrscheinlichkeit für das Fangen von s markierten Fischen.
Prinzip der Maximum-Likelihood-Schätzung:
Wähle n so, dass bei Ergebnis s die Wahrscheinlichkeit $h(s; k, n, S)$ für diese Realisation maximal wird.
Ergebnis: $\hat{n} = k \cdot \frac{S}{s}$.

Beispiel: $S = 1000$, $k = 1000$, $s = 100 \Rightarrow \hat{n} = 1000 \cdot \frac{1000}{100} = 10000$.
Fragen in der Statistik:
$\hat{n}$: „Schätzung“ in welchem Sinn?
Eigenschaften der Schätzung?
Besser: Intervall angeben, so dass „wahrer“ Wert mit vorgegebener Vertrauenswahrscheinlichkeit in diesem Intervall $\rightarrow$ Konfidenzintervall; u.s.w.

(2.15) **Beispiel**: Anwendung wie (2.14), aber *mit* Zurücklegen, mit Reihenfolge.
Anzahl der möglichen Stichproben: n^k.
Gesucht: Anzahl der Stichproben, die s defekte Teile enthalten.
Dazu: S^s Möglichkeiten, s defekte Teile aus S auszuwählen.
$(n-S)^{k-s}$ Möglichkeiten, $k - s$ intakte Teile aus $n - S$ auszuwählen.



Hier: Ziehen mit Zurücklegen *und* mit Reihenfolge.
Daher: Anzahl der Möglichkeiten, s defekte Teile auf k Plätze zu verteilen $\binom{k}{s}$.
Damit: Wahrscheinlichkeit für s defekte Teile ($s \in \{0, ..., k\}$):

$$b(s; k, \frac{S}{n}) = \frac{1}{n^k} \cdot \binom{k}{s} \cdot S^s \cdot (n-S)^{k-s} = \binom{k}{s} \left(\frac{S}{n}\right)^s \cdot \left(1 - \frac{S}{n}\right)^{k-s} = \binom{k}{s} \cdot p^s \cdot (1-p)^{k-s},$$

wobei $p := \frac{S}{n}$ (Schlechtanteil).
Die durch b bestimmte Zähldichte (als Funktion von s) definiert die sogenannte Binomialverteilung:
kurz: $b(s; k, p)$ oder $b(k, p)$.
(„Probe“: $\sum_{s=0}^{k} \binom{k}{s} \cdot p^s \cdot (1-p)^{k-s} \underset{\text{(Binomialsumme)}}{=} (p + (1-p))^k = 1$).
Intuitiv: n groß $\rightarrow$ kaum Unterschied zwischen Ziehen mit und ohne Zurücklegen.
Dazu:

(2.16) **Lemma**: Sei $(S_n)_{n \in \mathbb{N}}$ mit $\lim_{n \to \infty} \frac{S_n}{n} = p \in (0, 1)$.
Dann gilt: $\lim_{n \to \infty} h(s; k, n, S_n) = b(s; k, p)$.
Beweis: siehe Übung.
Siehe oben: Eine Produktion enthalte den Anteil p defekter Teile. Dann ist die Wahrscheinlichkeit für s defekte Teile in einer Stichprobe vom Umfang k (mit Zurücklegen) $b(s; k, p) = \binom{k}{s} \cdot p^s \cdot (1-p)^{k-s}$.

Was passiert bei $k \rightarrow \infty$ und $p \rightarrow 0$?

(2.17) **Lemma**: Gegeben sei Folge von Binomialverteilungen $b(s;k,p_k), k \in \mathbb{N}$ mit $k \cdot p_k = \lambda > 0\ \forall k \in \mathbb{N}$.
Dann gilt: $\lim_{k\to\infty} b(s;k,p_k) = \frac{\lambda^s}{s!}e^{-\lambda}$ für $s \in \mathbb{N}_0$.
Beweis:

$$\lim_{k\to\infty} \binom{k}{s} \cdot {p_k}^s \cdot (1-p_k)^{k-s} = \lim_{k\to\infty} \frac{k!}{s!(k-s)!} \cdot \left(\frac{\lambda}{k}\right)^s \cdot \left(1-\frac{\lambda}{k}\right)^{k-s}$$

$$= \lim_{k\to\infty} \frac{\lambda^s}{s!} \cdot \underbrace{\frac{k\cdot(k-1)\cdot\ldots\cdot(k-s+1)}{k^s}}_{\underbrace{1\cdot\frac{(k-1)}{k}}_{\to 1}\cdot\ldots\cdot\underbrace{\frac{k-s+1}{k}}_{\to 1}} \cdot \underbrace{\left(1-\frac{\lambda}{k}\right)^k}_{\to e^{-\lambda}} \cdot \underbrace{\left(1-\frac{\lambda}{k}\right)^{-s}}_{\to 1} = \frac{\lambda^s}{s!}\cdot e^{-\lambda}.$$

Offensichtlich definiert $p(s) := \frac{\lambda^s}{s!} \cdot e^{-\lambda}$ eine Zähldichte auf $\Omega = \mathbb{N}_0$.
Sei $\lambda > 0$ beliebig:

$$\sum_{s=0}^{\infty} p(s) = e^{-\lambda} \underbrace{\sum_{s=0}^{\infty} \frac{\lambda^s}{s!}}_{e^{\lambda}} = 1.$$

(2.18) **Bezeichnung**: Die Wahrscheinlichkeitsverteilung auf $\mathbb{N}_0$ definiert durch die Zähldichte $p(s) = \frac{\lambda^s}{s!} \cdot e^{-\lambda}, s \in \mathbb{N}_0, \lambda > 0$ heißt Poissonverteilung mit Parameter λ; kurz: $\mathrm{po}(s;\lambda)$ bzw. $\mathrm{po}(\lambda)$.

Bemerkung:

- (2.17) heißt auch Gesetz der seltenen Ereignisse (wegen $p_k \to 0$, $k \to \infty$).
  Beispiel: radioaktiver Zerfall, Häufigkeiten von Fehlern in Systemen, Anzahlen von Telefonaten, ... $\to$ Bartbiewicz 1898

- (2.17) gilt auch unter der Voraussetzung $k \cdot p_k \underset{k\to\infty}{\to} \lambda\ (> 0)$.

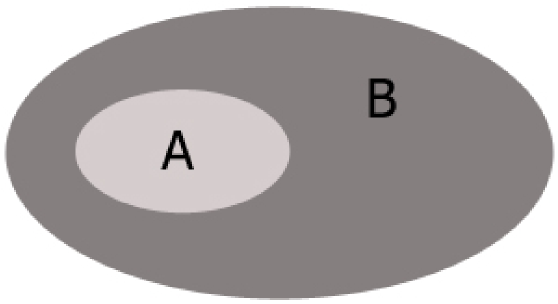
1.3 Eigenschaften von Wahrscheinlichkeitsräumen

Sei in diesem Abschnitt $(\Omega, \mathfrak{A}, P)$ ein allgemeiner Wahrscheinlichkeitsraum.

(3.1) **Lemma**: Es gelten für $A, B \in \mathfrak{A}$:

i) $P(A \cup B) = P(A) + P(B)$, falls $A \cap B = \emptyset$.

ii) $P(B \backslash A) = P(B) - P(A)$, falls $A \subset B$ (Subtraktivität von P).



iii) $P(A^C) = 1 - P(A)$

iv) $A \subset B \Rightarrow P(A) \leq P(B)$ (Monotonie von P)

v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

vi) $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{n} P(A_i)$, $A_1, A_2, ... \in \mathfrak{A}$.

Beweis:

i) klar aus σ-Additivität.

ii) $B \backslash A = B \cap A^C$
Es ist: $P(B) = P(B \cap \Omega) = P(B \cap (A \cup A^C)) = P(\underbrace{(\underbrace{B \cap A}_{A}) \cup (B \cap A^C)}_{\text{disjunkt}}) \overset{(i)}{=} P(A) + P(B \cap A^C)$.

iii) aus ii) mit $B = \Omega$.

iv) $P(A) = P(B) - P(B \backslash A) \leq P(B)$.

v) Wegen Additivität gilt:
$P(A^C \cap B) + P(A \cap B) = P(B)$
$P(A \cap B^C) + P(A \cap B) = P(A)$
Nun ist $A \cup B = \underbrace{(A \cap B^C) \cup (A \cap B) \cup (A^C \cap B)}_{\text{disjunkt}}$
$\Rightarrow P(A \cup B) = P(A \cap B^C) + P(A \cap B) + P(A^C \cap B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) = P(A) + P(B) - P(A \cap B)$.

vi) Trick: Ereignis diskunkt aufspalten und dann Additivität nutzen.
Es ist: $\bigcup_{i=1}^{n} A_i = A_1 \cup ({A_1}^C \cap A_2) \cup ({A_1}^C \cap {A_2}^C \cap A_3) \cup ... \cup (\bigcap_{j=1}^{n-1} {A_j}^C \cap A_n) = A_1 \cup \bigcup_{i=2}^{n} ((\bigcap_{j=1}^{n} {A_j}^C) \cap A_i)$
$\Rightarrow P(\bigcup_{i=1}^{n} A_i) = P(A_1) + \sum_{i=2}^{n} P((\bigcap_{j=1}^{i-1} {A_j}^C) \cap A_i)$
$\underset{(\text{iv})}{\leq} P(A_1) + \sum_{i=2}^{n} P(A_i) = \sum_{i=1}^{n} P(A_i)$.

(3.2) **Definition**: (Limites (Grenzwerte) von Ereignisfolgen)
Sei $(A_n)_{n\in\mathbb{N}} \subseteq \mathfrak{A}$, $\mathfrak{A}$ sei σ-Algebra über $\Omega \neq \emptyset$.
$(A_n)_n$ heißt monoton wachsend, falls $A_n \subseteq A_{n+1}$ $\forall n \in \mathbb{N}$.
$(A_n)_n$ heißt monoton fallend, falls $A_n \supseteq A_{n+1}$ $\forall n \in \mathbb{N}$.
(kurz: $(A_n)_n \uparrow$ bzw. $(A_n)_n \downarrow$.
Für monoton wachsende, bzw. fallende Ereignisfolgen heißt jeweils

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$$

bzw.

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$$

der Limes von $(A_n)_{n\in\mathbb{N}}$.
Für eine *beliebige* Ereignisfolge $(A_n)_n$ heißen

$$\limsup_{n\to\infty} A_n = \lim_{n\to\infty}(\bigcup_{k=n}^{\infty} A_k) = \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{k=n}^{\infty} A_k}_{\downarrow}$$

der limes superior und

$$\liminf_{n\to\infty} A_n = \lim_{n\to\infty}(\bigcap_{k=n}^{\infty} A_k) = \bigcup_{n=1}^{\infty} \underbrace{\bigcap_{k=n}^{\infty} A_k}_{\uparrow}$$

der limes inferior von $(A_n)_{n\in\mathbb{N}}$.

(3.3) **Bemerkung**: Es ist mit $(A_n)_n \subset \mathfrak{A}$:
$\limsup_{n\to\infty} A_n \in \mathfrak{A}$, $\liminf_{n\to\infty} A_n \in \mathfrak{A}$ (siehe (1.8) und (1.9)) und
$\limsup_{n\to\infty} A_n = \{\omega \in \Omega; \omega$ liegt in unendlich vielen der A_n's$\}$ heißt: unendlich viele der A_i treten ein.
$\liminf_{n\to\infty} A_n = \{\omega \in \Omega; \omega$ liegt in allen A_n's bis auf endlich viele$\}$ heißt: alle bis auf endlich viele der A_i treten ein (fast alle A_i's treten ein).
Beweis: siehe Übung.

(3.4) **Lemma**: $(\Omega, \mathfrak{A}, P)$ WR, $(A_n)_n \subseteq \mathfrak{A}$. Dann gelten:

i) $P(\bigcup_{n=1}^{\infty} A_n) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$, falls $(A_n)_n \uparrow$ (Stetigkeit von P von unten).

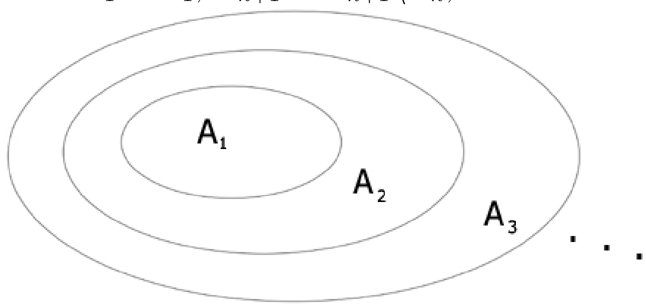
ii) $P(\bigcap_{n=1}^{\infty} A_n) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$, falls $(A_n)_n \downarrow$ (Stetigkeit von P von oben).

iii) $P(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} P(\bigcup_{k=n}^{\infty} A_k)$
$P(\liminf_{n\to\infty} A_n) = \lim_{n\to\infty} P(\bigcap_{k=n}^{\infty} A_k)$

iv) $P(\bigcup_{n=1}^\infty A_n) \leq \sum_{n=1}^\infty P(A_n)$ (Sub-σ-Additivität).

Beweis:

i) Seien $B_1 = A_1$, $B_{n+1} = A_{n+1} \backslash A_n$, $n \in \mathbb{N}$.



$\Rightarrow (B_n)_n$ paarweise diskunkt und
$\bigcup_{n=1}^\infty B_n = B_1 \cup \bigcup_{n=2}^\infty \underbrace{(A_n \cap A_{n-1}{}^C)}_{\subseteq A_n} \subseteq \bigcup_{n=1}^\infty A_n$
und $\omega \in \bigcup_{n=1}^\infty A_n \Rightarrow \exists i : \omega \in A_i$
und $\omega \notin A_j, j < i \Rightarrow \omega \in B_i \Rightarrow \omega \in \bigcup_{i=1}^\infty B_n$.
Also: $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n$.
Damit:
$P(\bigcup_{n=1}^\infty A_n) = P(\bigcup_{n=1}^\infty B_n) \overset{\sigma\text{-Add.}}{=} \sum_{n=1}^\infty P(B_n)$
$= P(A_1) + \sum_{n=2}^\infty P(B_n) = P(A_1) + \lim_{m\to\infty} \sum_{n=2}^m P(B_n)$
$= P(A_1) + \lim_{m\to\infty} \underbrace{\sum_{n=1}^m \underbrace{P(B_{n+1})}_{P(A_{n+1})-P(A_n)}}_{P(A_{m+1})-P(A_1)} = \lim_{m\to\infty} P(A_m)$.

ii) Mit de Morgan aus i):
$\bigcap_{n=1}^\infty A_n = (\bigcup_{n=1}^\infty A_n{}^C)^C$, $(A_n{}^C)_n \uparrow$
$\Rightarrow P(\bigcap_{n=1}^\infty A_n) = 1 - P(\bigcup_{n=1}^\infty A_n{}^C) \overset{i)}{=} 1 - \lim_{m\to\infty} P(A_m{}^C) = \lim_{m\to\infty} P(A_m)$

iii) $P(\limsup_{n\to\infty} A_n) = P(\lim_{n\to\infty} \bigcup_{k=n}^\infty A_k) = \lim_{n\to\infty} P(\bigcup_{k=n}^\infty A_k)$
$P(\liminf_{n\to\infty} A_n) = P(\lim_{n\to\infty} \bigcap_{k=n}^\infty A_k) = \lim_{n\to\infty} P(\bigcap_{k=n}^\infty A_k)$

iv) analog zu (3.1) vi) mit Verwendung der σ-Additivität (in (3.1) für endlich viele Ereignisse).

(3.5) **Lemma**: (Siebformel von Sylvester Poincaré)
Für Ereignisse $(A_n)_{n\in\mathbb{N}}$ in einem WR $(\Omega, \mathfrak{A}, P)$ gilt:

$$P(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n P(A_k) - \sum_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2})$$
$$+ \sum_{1\leq i_1<i_2<i_3\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3})$$

$$\mp ... + (-1)^{n+1} \cdot P(\bigcap_{k=1}^{n} A_k).$$

Speziell: $n = 2$: $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

Struktur von (3.5) aus der Kombinatorik bekannt: Einschluss-Ausschluss-Prinzip.
(3.5) kann zur Lösung kombinatorischer Aufgaben verwendet werden.

(3.6) **Beispiel**: Sei $\Omega_n = \{\pi;\ \pi : \{1, ..., n\} \to \{1, ..., n\} \text{ bijektiv}\}$, das heißt, Ω_n ist die Menge aller Permutationen aus n Dingen.
$\Rightarrow$ ((2.5), $k = n$) $|\Omega_n| = (n)_n = n!$.
Seien $A = \{\pi \in \Omega_n;\ \pi(i) \neq i,\ 1 \leq i \leq n\}$ die Menge der fixpunktfreien Permutationen und P die Laplaceverteilung auf Ω_n.
Weiterhin: $A_i := \{\pi \in \Omega_n;\ \pi(i) = i\}$, $1 \leq i \leq n$ (Fixpunkt an Stelle i).
$\Rightarrow A = (\bigcup_{i=1}^n A_i)^C$
$\Rightarrow P(A) = 1 - P(\bigcup_{i=1}^\infty A_i)$
$\overset{(3.5)}{=} 1 - \sum_{i=1}^n P(A_i) + \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \mp ... + (-1)^{n+1} \cdot P(\bigcap_{k=1}^n A_k)$
Nun ist:
$|A_i| = (n-1)!$ (eine Stelle fest, alle anderen permutiert)
$i_1 < i_2$: $|A_{i_1} \cap A_{i_2}| = (n-2)!$ (zwei Stellen fest, alle anderen permutiert)
usw., das heißt, $|A_{i_1} \cap ... \cap A_{i_k}| = (n-k)!$, $i_1 < ... < i_k$
$\Rightarrow P(A) = 1 - \frac{1}{n!} \cdot \sum_{k=1}^n (-1)^{k+1} \cdot \sum_{1 \leq i_1 < ... < i_k \leq n} (n-k)!$
$= 1 - \sum_{k=1}^n (-1)^{k+1} \cdot \frac{(n-k)!}{n!} \cdot \binom{n}{k}$
$= \sum_{k=0}^n (-1)^k \cdot \frac{1}{k!}$.
Weiterhin: $B_{n,m} = \{\pi \in \Omega_n;\ \pi \text{ hat genau } m \text{ Fixpunkte}\}$
$= \{\pi \in \Omega_n;\ \pi(i_j) = i_j,\ 1 \leq j \leq m,\ i_j \neq i_l \text{ für } j \neq l,\ \pi(i) \neq i\ \forall i \neq i_j\}$.
$\Rightarrow P(B_{n,m}) = \frac{1}{m!} \cdot \sum_{k=0}^{n-m} (-1)^k \cdot \frac{1}{k!}$, $(A \equiv B_{n,0})$
(denn: genau m Fixpunkte $\to \binom{n}{m}$ Möglichkeiten der Auswahl von m Fixpunkten; unter den übrigen $n - m$ Stellen *keine* Fixpunkte (s.o.);

$$P(B_{n,m}) = \underbrace{\frac{1}{n!}}_{n! = |\Omega_n|} \cdot \binom{n}{m} \cdot \underbrace{(n-m)! \cdot \sum_{k=0}^{n-m} (-1)^k \cdot \frac{1}{k!}}_{\text{s.o.: bestimmte Anzahl}}.$$

Bemerkung: Interpretation über Sortierprobleme.
Gegeben Feld der Länge n.
Gesucht: Wahrscheinlichkeit, dass mindestens $k \leq n$ Elemente des Feldes schon an der richtigen Stelle stehen, wenn die Elemente bezüglich eines ordinalen Merkmals sortiert werden sollen.
Voraussetzung: Laplaceraum: Jede der $n!$ möglichen Anordnungen hat Wahrscheinlichkeit $\frac{1}{n!}$.

- Wahrscheinlichkeit, dass mindestens ein Element an der richtigen Stelle: $1 - P(A)$.
- Wahrscheinlichkeit, dass in einem Feld der Länge n bereits k Elemente richtig sortiert sind: $P(B_{n,k})$.

– A_i: Menge aller Eingabefolgen mit der richtigen Sortierung des i-ten Elements.

(3.7) **Bemerkung**: Es gilt:

$$|\sum_{k=0}^{n-m}(-1)^k \cdot \frac{1}{k!} - e^{-1}| = |\sum_{k=0}^{n-m}(-1)^k \cdot \frac{1}{k!} - \sum_{k=0}^{\infty}(-1)^k \cdot \frac{1}{k!}|$$
$$= |\sum_{k=n-m+1}^{\infty}(-1)^k \cdot \frac{1}{k!}| \leq \frac{1}{(n-m+1)!},$$

da alternierende Reihe.
Unterscheidung:

– $n \geq 8$ und $m \geq 5 \rightarrow \frac{1}{m!} \leq \frac{1}{5!}$
– $n \geq 8$ und $m \leq 4 \rightarrow \frac{1}{(n-m+1)!} \leq \frac{1}{5!}$

$\Rightarrow$ Ist $n \geq 8$ und m beliebig, dann ist

$$|P(B_{n,m}) - \frac{e^{-1}}{m!}| \leq \frac{1}{m!} \cdot \frac{1}{(n-m+1)!} \leq \frac{1}{5!} < 0,01$$

(3.8) **Beispiel**: n Personen geben Hut an Gaderobe ab, Rückgabe zufällig:
Gesucht: Wahrscheinlichkeit, dass niemand den eigenen Hut erhält.
$\hookrightarrow P(B_{n,0}) = P(A) = \sum_{k=0}^{n}(-1)^k \cdot \frac{1}{k!} \approx \frac{e^{-1}}{0!} \approx 0,37$
mit einem Fehler von $\leq 0,01$; unabhängig von n.
Folgerung aus Sylvester-Poincaré.

(3.9) **Korollar** (Bonferroni-Ungleichungen)
Seien $A_1, ..., A_n \in \mathfrak{A}$ in WR $(\Omega, \mathfrak{A}, P)$. Dann gilt:

$$\sum_{k=1}^{n} P(A_k) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \leq P(\bigcup_{k=1}^{n} A_k) \underset{(3.1)}{\leq} \sum_{k=1}^{n} P(A_k).$$

Weitere Bonferroni-Ungleichungen entstehen durch Abbruch der Siebformel nach Termen gerader oder ungerader Ordnung.

1.4 Bedingte Wahrscheinlichkeit

Idee: Verarbeitung von Vorinformation, bzw. Zusatzinformation.
Für $A \in \mathfrak{A}$ ist $P(A)$ Wahrscheinlichkeit des Eintretens; jetzt bekannt / gefordert:
$B \in \mathfrak{A}$ tritt ein.
$\hookrightarrow$ Einfluss auf $P(A)$?

(4.1) **Beispiel**:

i) Würfel: Frage: Wahrscheinlichkeit für die 2 unter der Bedingung, dass eine gerade Zahl auftritt.
$\hookrightarrow P(\{2\}|\{2,4,6\}) = \frac{1}{3}$
also: WR wird eingeschränkt.

ii) Ziehen von Kugeln ohne Zurücklegen.
Urne mit 2 weißen und 3 schwarzen Kugeln, Ziehen von 2 Kugeln.
P(2. Kugel schwarz | 1. Kugel ist weiß) $= \frac{3}{4} \hookrightarrow$ allgemein?
Hier Laplace-Experiment: $|\Omega| = 5 \cdot 4$
weiße Kugeln: Nr. $1, 2$
schwarze Kugeln: Nr. $3, 4, 5$
Ereignis A: 2. Kugel ist schwarz.
Ereignis B: 1. Kugel ist weiß.
$A \cap B = \{(1,3),(1,4),(1,5),(2,3),(2,4),(2,5)\}, |A \cap B| = 6.$
$B = \{(1,2),(1,3),(1,4),(1,5),(2,1),(2,3),(2,4),(2,5)\}, |B| = 8.$
$P(A \cap B) = \frac{6}{20}$, $P(B) = \frac{8}{20}$.
$\Rightarrow \frac{3}{4} = P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{\frac{6}{20}}{\frac{8}{20}}$.

(4.2) **Beispiel**: relative Häufigkeiten
Befragung von 100 Personen.

Präferenz zu Produkt A oder B?

	weiblich	männlich	
A	10	20	30
B	50	20	70
	60	40	100

Tabelle der relativen Häufigkeiten:

	weiblich	männlich	
A	0,1	0,2	0,3
B	0,5	0,2	0,7
	0,6	0,4	1,0

relative Häufigkeit für B in der Gruppe der Frauen:
$\frac{\text{rel. Häufigkeit für } B \text{ und Frau}}{\text{rel. Häufigkeit für Frau}} = \frac{0{,}5}{0{,}6} = \frac{5}{6}$.

(4.3) **Definition**: Sei $(\Omega, \mathfrak{A}, P)$ ein WR.
Für jedes $B \in \mathfrak{A}$ mit $P(B) > 0$ wird durch $P(A|B) = \frac{P(A\cap B)}{P(B)}$, $A \in \mathfrak{A}$
eine Wahrscheinlichkeitsverteilung $P(\cdot|B)$ auf $\mathfrak{A}$ definiert, die sogenannte *bedingte Verteilung* unter (der Hypothese) B.
$P(A|B)$ heißt elementar bedingte Wahrscheinlichkeit von A und B.

Bemerkung: $(\Omega, \mathfrak{A}, P(\cdot|B))$ ist ein WR,
weil $P(A|B) = P(A \cap B|B)$, $(B, \underbrace{\{A \cap B; A \in \mathfrak{A}\}}_{\text{Spur-}\sigma\text{-Algebra}}, P(\cdot|B))$ auch ein WR ist,
der sogenannte induzierte oder eingeschränkte WR.
Zur Wohldefiniertheit: $P(\cdot|B)$ ist WV, denn:

i) $P(A|B) \geq 0$,

ii) $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = 1$,

iii) $P(\bigcup_{i=1}^{\infty} A_i|B) = \frac{P(\bigcup A_i \cap B)}{P(B)} = \frac{P(\bigcup (A_i \cap B))}{P(B)} = \frac{\sum P(A_i \cap B)}{P(B)} = \sum \underbrace{\frac{P(A_i \cap B)}{P(B)}}_{P(A_i|B)}$,

wobei $(A_i)_i$ paarweise disjunkt.

(4.4) **Lemma**: $A, B, A_1, A_2, ... \in \mathfrak{A}$
$P(A) > 0, P(B) > 0, P(\bigcup_{i=1}^{n-1} A_i) > 0$
Dann:

i) $P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}$

ii) $P(\bigcap_{i=1}^{n} A_i) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot ... \cdot P(A_n| \bigcap_{i=1}^{n-1} A_i)$

Beweis:

i) $P(B|A) \cdot \frac{P(A)}{P(B)} = \frac{P(B \cap A)}{P(A)} \cdot \frac{P(A)}{P(B)} = P(A|B)$

ii) $P(A_1) \cdot P(A_2|A_1) \cdot ... \cdot P(A_n| \bigcap_{i=1}^{n-1} A_i)$
$= P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot ... \cdot \frac{P(\bigcap_{i=1}^{n} A_i)}{P(\bigcap_{i=1}^{n-1} A_i)}$

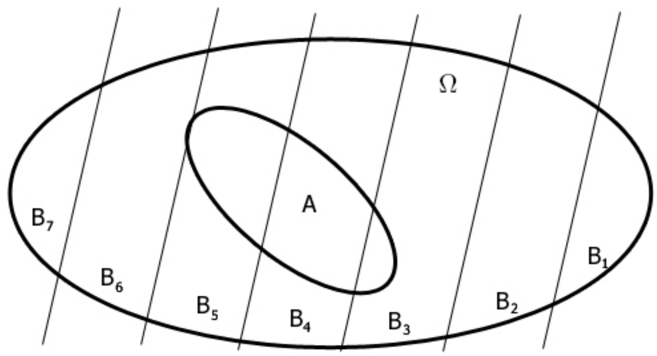
(4.5) **Beispiel**: Wahrscheinlichkeit, dass die drei Spieler je genau ein As haben.
A_i: Spieler i erhält genau ein As.
$\Rightarrow P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$
Modell:
10 Karten an Spieler 1, dann
10 Karten an Spieler 2, dann
10 Karten an Spieler 3, dann
2 Karten in den Skat (o.B.d.A. aus Symmetriegründen).
$P(A_1) = \frac{\binom{4}{1} \cdot \binom{28}{9}}{\binom{32}{10}}$ ($\rightarrow$ hypergeometrische Verteilung).
$P(A_2|A_1) = \frac{\binom{3}{1} \cdot \binom{19}{9}}{\binom{22}{10}}$
$P(A_3|A_1 \cap A_2) = \frac{\binom{2}{1} \cdot \binom{10}{9}}{\binom{12}{10}} \hookrightarrow$ Lösung.

(4.6) **Lemma**: (Formel der lokalen Wahrscheinlichkeit)
Gegeben: $A \in \mathfrak{A}, (B_n)_n \subset \mathfrak{A}, B_n$ paarweise disjunkt und $A \subset \bigcup_{n=1}^{\infty} B_n$

(häufig: $\bigcup B_n = \Omega$, das heißt, disjunkte Zerlegung von Ω)
Dann gilt:

$$P(A) = \sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n).$$

Ist $P(B_k) = 0$, so ist zunächst $P(A|B_k)$ nicht definiert, setze dann: $P(B_k) \cdot P(A|B_k) = 0$.



Beweis:
Wegen $A \subset \bigcup_{n=1}^{\infty} B_n$ ist $A = \bigcup_{n=1}^{\infty} (\underbrace{A \cap B_n}_{\text{disjunkt}})$

$P(A) \underset{\sigma\text{-Add.}}{=} \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n)$

(4.7) **Beispiel**:

i) Serienartikel wird parallel auf drei Fertigungsanlagen produziert → gemeinsames Transportband.
Megenanteile der drei Anlagen werden (etwa durch unregellmäßige Ausfallzeiten) als Wahrscheinlichkeit angesetzt:
$P(A_1) = 0,3,\ P(A_2) = 0,2,\ P(A_3) = 0,5$ mit
A_i: „Artikel wurde durch Anlage i hergestellt“
Weiterhin:
Wahrscheinlichkeit für fehlerhaften Artikel bei Anlage:

1	2	3
0,05	0,03	0,09

Gesucht:
Wahrscheinlichkeit, dass ein zufällig geprüftes Stück fehlerhaft ist?
Dazu: B: „Stück ist fehlerhaft“
Voraussetzung:
$P(B|A_1) = 0,05,\ P(B|A_2) = 0,03,\ P(B|A_3) = 0,09$
$\Rightarrow P(B) \overset{(4.6)}{=} \sum_{i=1}^{3} P(B|A_i) \cdot P(A_i) = 0,066.$

ii) Eine unter 1000000 Münzen hat Zahl auf beiden Seiten; alle Anderen sind „fair“ → Symbol / Zahl: jede Seite fällt mit Wahrscheinlichkeit $\frac{1}{2}$.
Eine zufällig ausgewählte Münze wird 20 mal geworfen. Ergebnis: 20 mal Zahl.
Frage:
Wahrscheinlichkeit, dass Münze trotzdem fair.
Dazu:

A: faire Münze wird gezogen.
B: unfaire Münze wird gezogen ($= A^C$).
$P(A) = \frac{10^6-1}{10^6} = 1 - 10^{-6}$, $P(B) = 10^{-6}$
z_{20}: „Es fällt 20 mal Zahl."
$P(z_{20}) = \underbrace{P(z_{20}|A)}_{\frac{1}{2^{20}}} \cdot P(A) + P(z_{20}|B) \cdot P(B)$ (lokale Wahrscheinlichkeit)
$= 2^{-20} \cdot (1 - 10^{-6}) + 10^{-6}$
$\overset{(4.4)(i)}{\Rightarrow} P(A|z_{20}) = P(z_{20}|A) \cdot \frac{P(A)}{P(z_{20})} = 0,4881.$

(4.8) **Satz**: (Bayes'sche Formel)
Seien $A, (B_n)_n \subset \mathfrak{A}, B_n$ paarweise disjunkt, $A \subset \bigcup_{n=1}^{\infty} B_n$ und $P(A) > 0$.
Dann gilt:
$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n)}$ $\forall k \in \mathbb{N}$.
Beweis:
$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{P(A)}$
$\Rightarrow$ Beh. mit (4.6), ‚Formel der lokalen Wahrscheinlichkeit'
Bayes'sche Formel:
Hier wird von „Wirkung" auf „Ursache" B_k zurückgeschlossen.

Beispiel:
Arzt stellt Symptom A fest, dass von verschiedenen Krankheiten $B_1, ..., B_n$ herrühren kann.

- relative Häufigkeit einer Krankheit bekannt $\rightarrow P(B_i)$
- Wenn Krankheit B_k vorliegt, dann relative Häufigkeit für das Auftreten von Symptom A bekannt $\rightarrow P(A|B_k)$.
  Gesucht:
  Wenn Symptom A auftritt, wie groß ist dann die Wahrscheinlichkeit für Krankheit B_k?
  Bemerkung:
  $P(B_k)$ heißen a priori.
  $P(B_k|A)$ heißen a posteriori.

(4.9) **Beispiel**:

i) Fortsetzung von (4.7)(i)
Gesucht:
Wahrscheinlichkeit, dass ein geprüftes und ein einwandfreies Stück auf Anlage 3 hergesellt wurde.

$$P(A_3|B^C) = \frac{P(A_3) \cdot P(B^C|A_3)}{\sum_{i=1}^{3} P(A_i) \cdot P(B^C|A_i)} = 0,49$$

$$P(B^C|A_1) = \frac{P(B^C \cap A_1)}{P(A_1)} = \frac{P(A_1) - P(B \cap A_1)}{P(A_1)}$$

ii) medizinisches Diagnoseverfahren für Krankheit

* in 90 Prozent der Fälle wird ein Kranker als „krank“ erkannt (Verfahren liefert richtigerweise positiven Befunde)
* in 5 Prozent der Fälle wird ein Gesunder als krank eingestuft (d.h., positiver, aber falscher Befund)

Modell:
1 % der Bevölkerung leidet an der Krankheit. Bestimme Wahrscheinlichkeit, dass eine zufällig ausgewühlte Testperson gesund ist, falls Diagnose positiven Befund liefert.
Lösung:
G: Person gesund.
B: Verfahren liefert positiven Befund.
Voraussetzung:
$P(B|G^C) = 0,9,\ P(B|G) = 0,05,\ P(G^C)0,01.$
$\Rightarrow P(B) = P(B|G^C) \cdot P(G^C) + P(B|G) \cdot P(G)$
$= 0,9 \cdot 0,01 + 0,05 \cdot 0,99 = 0,0585.$
Also:
$P(G|B) = \frac{P(B|G) \cdot P(G)}{P(B)} = 0,846$
(Aber: $P(B|G^C) = 0,9,\ P(B|G) = 0,01\ (0,001)$
$\Rightarrow P(G|B) = 0,52\ (0,10)$)

(4.10) **Beispiel**: (Gestörter Nachrichtenkanal)
Gesendet:
0 oder 1 über n Stationen
$S_0 \to S_1 \to ... \to S_n$
Voraussetzung:
Wahrscheinlichkeit für korrekte Übermittlung von S_i nach S_{i+1},
$i = 1, 2, ..., p \in (0, 1).$
Gesucht:
Wahrscheinlichkeit p_n, dass S_n die von S_0 gesendete Nachricht erhält.
Modell:
$\Omega_n = \{(\omega_0, ..., \omega_{n-1}); \omega_j \in \{0, 1\}, 0 \leq j \leq n-1\}$
(ω_i: Nachricht von S_i an S_{i+1})
Sei $E_m := \{\omega \in \Omega_n; \omega_0 = \omega_{m-1}, 1 \leq m \leq n\}$ (S_m erhält korrekte Nachricht)
Mit $p_m := P(E_m), 1 \leq m \leq n$, ist:
Voraussetzung:
$p_1 = 1$
$p_n = P(E_n) = P(E_{n-1} \cap E_n) + (E_{n-1}^C \cap E_n)$
$\underbrace{P(E_n|E_{n-1})}_{p} \cdot \underbrace{P(E_{n-1})}_{p_{n-1}} + \underbrace{P(E_n|E_{n-1}^C)}_{1-p} \cdot \underbrace{P(E_{n-1}^C)}_{1-p_{m-1}}$
$= (2p-1) \cdot p_{n-1} + 1 - p', n \geq 2$
$\Rightarrow$ (Induktion) $p_n = \frac{1}{2} + \frac{1}{2} \cdot (2p-1)^{n-1}$
$\Rightarrow \lim_{n \to \infty} p_n = \frac{1}{2}\ (|2p-1| < 1)$

p\n	10	100
0,9	0,57	0,5
0,99	0,92	0,57
0,999	0,99	0,91

1.5 Stochastische Unabhängigkeit

Heuristisch:
Zwei Ereignisse A und B sind unabhängig, falls $P(A)$ nicht von der Kenntnis des Eintretens oder Nicht-Eintretens von B abhängt, d.h., $P(A|B) = P(A)$, bzw. $P(B|A) = P(B)$.

Beispiel:
Urne mit 2 weißen und 3 schwarzen Kugeln
$\hookrightarrow$ Ziehen von 2 Kugeln mit Zurücklegen
$\Omega = \{(i,j); 1 \leq i,j \leq 5\}, |\Omega| = 5^2$
weiße Kugeln Nr. $1, 2$
schwarze Kugeln Nr. $3, 4, 5$
A: „2. Kugel schwarz“ $\hookrightarrow A = \{(i,j) \in \Omega; j \in \{3,4,5\}\}$ mit $|A| = 5 \cdot 3$.
B: „1. Kugel weiß“ $\hookrightarrow B = \{(i,j) \in \Omega; i \in \{1,2\}\}$ mit $|B| = 2 \cdot 5$.
$\Rightarrow P(A) = \frac{15}{25} = \frac{3}{5}$, $P(B) = \frac{10}{25} = \frac{2}{5}$
$P(A \cap B) = \frac{6}{25}$
$P(A|B) = \frac{\frac{6}{25}}{\frac{2}{5}} = \frac{3}{5} = P(A)$
Klar, denn wegen Zurücklegen beeinflussen sich die Ziehungen nicht.
Weiterhin:
$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$ (I)
$P(A \cap B) = P(A) \cdot P(B)$ (II)
Definition über (I): $P(B) > 0$ als Voraussetzung
Definition über (II): keine Voraussetzung

(5.1) **Definition**:
Gegeben WR $(\Omega, \mathfrak{A}, P)$. $A, B \in \mathfrak{A}$ heißen stochastisch unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$.

(5.2) **Lemma**:

i) Mit A, B sind auch A, B^C und A^C, B^C stochastisch unabhängig.
ii) Ist $P(B) > 0$, so gilt:
A, B stochastisch unabhängig $\Leftrightarrow P(A|B) = P(A)$.
iii) Ist A eine Nullmenge, d.h., $P(A) = 0$, so sind A, B stochastisch unabhängig $\forall B \in \mathfrak{A}$.

Beweis:

i) Voraussetzung:
A, B stochastisch unabhängig.
$P(A \cap B^C) = P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) = P(A) \cdot (1 - P(B)) = P(A) \cdot P(B^C)$
$P(A^C \cap B^C) = P((A \cup B)^C) = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B)) = 1 - (P(A) + P(B) - P(A) \cdot P(B)) = (1 - P(A)) \cdot (1 - P(B))$
ii) siehe oben

iii) $P(A) = 0 \Rightarrow P(A \cap B) \leq P(A) = 0 \Rightarrow P(A \cap B) = 0$

(5.3) **Bemerkung**:
Unabhängigkeit abhängig von WV.
Sei $\Omega = \{1, 2, 3\}$, $A = \{1\}$, $B = \{1, 2\}$
$P = \varepsilon_1$, Q: Laplaceverteilung
$\Rightarrow \varepsilon_1(A \cap B) = \varepsilon_1(1) = 1 = \varepsilon_1(A) \cdot \varepsilon_1(B)$,
aber $Q(A \cap B) = \frac{1}{3} \neq \frac{2}{9} = \frac{1}{3} \cdot \frac{2}{3} = Q(A) \cdot Q(B)$

Nun allgemeinere Definition für 2 und mehr Ereignisse.

(5.4) **Definition**:
Familie $(A_i)_{i \in I} \subset \mathfrak{A}$, heißt paarweise stochastisch unabhängig, falls

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \; \forall i, j \in I, i \neq j.$$

(5.5) **Beispiel**:
Werfen von zwei unverfälschten Würfeln, d.h.,
Laplaceverteilung über $\Omega = \{(i, j); i, j \in \{1, ..., 6\}\}$.
$A_i \hat{=}$ Würfel i zeigt gerade Zahl, $i = 1, 2$
$\Rightarrow P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = \frac{1}{4}$, d.h., A_1, A_2 sind stoch. unabhängig.
Weiterhin $A_3 \hat{=}$ Summe der Augenzahl ist gerade
$\Rightarrow P(A_3|A_1) = P(A_3)$ und $P(A_3|A_2) = P(A_3)$, d.h., $\{A_1, A_2, A_3\}$ paarweise unabhängig.
Aber $P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3)$, denn $P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2) = \frac{1}{4}$.
$P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{1}{8}$.

Stärkerer Begriff der stochastischen Unabhängigkeit nötig.

(5.6) **Definition**:
Eine Familie $(A_i)_{i \in I}$ heißt stochastisch unabhängig (auch vollständig stochastisch unabhängig), falls für jede endliche Auswahl von Ereignissen gilt:

$$P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j) \; \forall J \neq \emptyset, J \subseteq I, J < \infty.$$

Bemerkung:

i) $(A_i)_i$ stoch. unabhängig $\Rightarrow$ $(A_i)_i$ paarweise stoch. unabhängig. Umkehrung gilt nicht, siehe Beispiel (5.5).

ii) Jede Teilfamilie einer stoch. unabhängigen Familie ist stoch. unabhängig.

iii) Beachte: (5.6) liefert System von Gleichungen:
z.B. A, B, C stoch. unabhängig, falls
$P(A \cap B) = P(A) \cdot P(B)$
$P(A \cap C) = P(A) \cdot P(C)$
$P(B \cap C) = P(B) \cdot P(C)$ und
$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$.

(5.7) **Satz**:

i) Seien $(A_i)_{i\in I}$ eine Familie stoch. unabhängiger Ereignisse, $k \notin I$ und $P(A_k) \in \{0,1\}$.
$\Rightarrow (A_i)_{i\in I\cup\{k\}}$ stoch. unabhängig.

ii) $(A_i)_{i\in I}$ stoch. unabhängig und $B_i \in \{A_i, A_i^C, \emptyset, \Omega\}$ $\forall i \in I \Rightarrow (B_i)_{i\in I}$ stoch. unabhängig.

iii) Sei $I = \{1, ..., n\}, n \in \mathbb{N}$ fest gewählt.
$(A_i)_{i\in I}$ stoch. unabhängig
$\Leftrightarrow P(\bigcap_{j=1}^n B_i) = \prod_{i=1}^n P(B_i)$ $\forall B_i \in \{A_i, A_i^C\}$ $\forall i \in \mathbb{N}$.

(5.8) **Beispiel**:
Experiment liefert mit Wahrscheinlichkeit p ein Ereignis A, mit $q = 1-p$ Ereignis A^C.
Experiment wird n-mal „unabhängig“ ausgeführt $\rightarrow$ Modell $\Omega = \{(\omega_1, ..., \omega_n); \omega_i \in \{\underbrace{0}_{\widehat{=}A^C}, \underbrace{1}_{\widehat{=}A}\}, 1 \leq i \leq n\}$.
Interpretiert man experimentelle Unabhängigkeit als stochastische Unabhängigkeit:
$\Rightarrow$ jedes ω mit k Komponenten gleich Eins Wahrscheinlichkeit $p(\omega) = p^k \cdot (1-p)^{n-k}$.
Wegen $\sum_{\omega\in\Omega} p(\omega) = \sum_{k=0}^n \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = 1$
ist (Ω, p) ein endlicher WR. Dieses Modell heißt *Bernoulli-Modell*.

(5.9) **Bemerkung**:
Sei die Familie $(A_i)_{i\in\{1,...,n\}}$ stoch. unabhängig oder kurz $A_1, ..., A_n$ stoch. unabhängig, dann gilt:

$$P(\bigcup_{i=1}^n A_i) = 1 - P(\bigcap_{i=1}^n A_i^C) \overset{(5.2)}{=} 1 - \prod_{i=1}^n P(A_i^C) = 1 - \prod_{i=1}^n (1 - P(A_i)).$$

(5.10) **Beispiel**:
Das Ziegen-Problem (aus der Show „Let's make a deal“).
Situation:

- 3 Türen, dahinter 1 Auto, 2 Ziegen
- Kandidat wählt
- Eingriff des Quizmasters
  Er öffnet eine der beiden nicht gewählten Türen, hinter der eine Ziege steht und erlaubt dem Kandidaten, seine Entscheidung zu ändern.

Frage: Änderung der Entscheidung oder nicht?
Konkret (o.B.d.A. wegen Symmetrie):
Kandidat wählt Tür 1, Quizmaster öffnet Tür 3.
Soll Kandidat bei Tür 1 bleiben oder Tür 2 wählen?
Dazu (ohne spezifizierten WR):

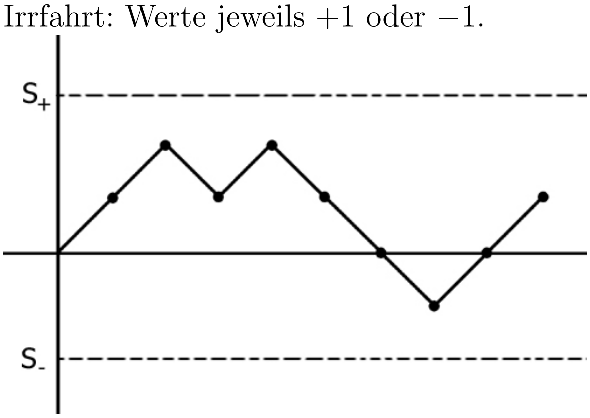
$A_i \hat{=}$ Auto hinter Tür i, Vor.: $P(A_i) = \frac{1}{3}$.
$K_i \hat{=}$ Kandidat wählt Tür i, Vor.: $P(K_i) = \frac{1}{3}$.
A_i, K_j unabhängig, $1 \leq i, j \leq 3$.
$Q_i \hat{=}$ Quizmaster öffnet Tür i (nicht unabhängig von A_i, K_j).
Rechnung:
$P(A_1|K_1 \cap Q_3) = \frac{P(A_1 \cap K_1 \cap Q_3)}{P(K_1 \cap Q_3)} = \frac{P(Q_3|A_1 \cap K_1)}{P(K_1 \cap Q_3)} \cdot P(A_1 \cap K_1)$ $(*)$
Es ist: $P(K_1 \cap Q_3) = P(K_1 \cap Q_3 \cap A_1) + P(K_1 \cap Q_3 \cap A_2) + \underbrace{P(K_1 \cap Q_3 \cap A_3)}_{=0}$

$\underbrace{= P(Q_3|A_1 \cap K_1)}_{\text{zusätzl. Vor. } \to \frac{1}{2}} \cdot \underbrace{P(A_1 \cap K_1)}_{\frac{1}{9}} + \underbrace{P(Q_3|A_2 \cap K_1)}_{1} \cdot \underbrace{P(A_2 \cap K_1)}_{\frac{1}{9} \text{ weg. unabh.}}$
$= \frac{1}{2} \cdot \frac{1}{9} + 1 \cdot \frac{1}{9} = \frac{1}{6}$
$\Rightarrow (*) = \frac{\frac{1}{2}}{\frac{1}{6}} \cdot \frac{1}{3} = \frac{1}{3}$.
Aber: $P(A_2|K_1 \cap Q_3) \overset{\text{analog}}{=} \frac{2}{3}$.
D.h., die Änderung der Entscheidung verdoppelt die Gewinnwahrscheinlichkeit!

(5.11) **Bemerkung**:
Relation „stoch. unabhängig“ ist nicht transitiv, d.h., aus A_1, A_2 und A_2, A_3 stoch. unabhängig folgt nicht notwendigerweise, dass A_1, A_3 stoch. unabhängig.
$\hookrightarrow$ WR $(\Omega, \mathfrak{A}, P)$ mit $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$, P: Laplace-Verteilung.
$A_1 = \{(0,0), (0,1)\}$, $A_2 = \{(0,1), (1,0)\}$, $A_3 = \{(1,0), (1,1)\}$
$\Rightarrow P(A_i) = \frac{1}{2}$, $i = 1, 2, 3$ und
$P(A_1 \cap A_2) = P((0,1)) = \frac{1}{4} = P(A_1) \cdot P(A_2)$
$P(A_2 \cap A_3) = P((1,0)) = \frac{1}{4} = P(A_2) \cdot P(A_3)$, aber
$P(A_1 \cap A_3) = P(\emptyset) = 0 \neq \frac{1}{4} = P(A_1) \cdot P(A_3)$.

Folgen von stoch. unabhängigen Ereignissen spielen eine zentrale Rolle in der Stochastik.

- Simulation
- Münzwurf, Frage: Wann fällt zum ersten Mal Zahl?
- diskrete Vert. mit $P(\{0\}) > 0$
  Frage: Wann ist Summe der Ereignisse zum ersten Mal größer als S?
- Irrfahrt: Werte jeweils $+1$ oder -1.



Fragen: Wann überschreitet Summenpfad erstmals die Grenze S_+?

(5.12) **Wiederholung**:
Bestimmung von Wahrscheinlichkeiten für $\limsup A_n$, $\liminf A_n$, $(A_n)_n \subset \mathfrak{A}$ Folge von Ereignissen

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k,\ \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k.$$

(5.13) **Bemerkung**:
Sei $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{A}$.
$(A_n)_n$ heißt konvergent $\Leftrightarrow \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$.
Es gilt stets: $\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$, denn:
Sei $\omega \in \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k \Rightarrow \exists n_0 : \omega \in \bigcap_{k=n_0}^{\infty} A_k$
$\Rightarrow \omega \in A_k\ \forall k \geq n_0$
$\Rightarrow \omega \in \bigcup_{k=n}^{\infty} A_k\ \forall n \geq 1 \Rightarrow \omega \in \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$
Weiterhin

$$(\limsup_{n\to\infty} A_n)^C = (\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k)^C = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^C = \liminf_{n\to\infty} A_n^C.$$

(5.14) **Lemma** (von Borch-Cantelli):
Seien $(\Omega, \mathfrak{A}, P)$ ein WR und $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{A}$.
Dann gilt:

i) $\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(\limsup_{n\to\infty} A_n) = 0$.
ii) Ist zusätzlich $(A_n)_n$ stoch. unabhängig:
$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\limsup_{n\to\infty} A_n) = 1$.

Beweis:

i) Wegen $\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k \subset \bigcup_{k=n}^{\infty} A_k\ \forall n \in \mathbb{N}$
ist $P(\limsup_{n\to\infty} A_n) \leq P(\bigcup_{k=n}^{\infty} A_k) \leq \sum_{k=n}^{\infty} P(A_k) \underset{n\to\infty}{\to} 0$.
ii) Es ist $P(\limsup_{n\to\infty} A_n) \overset{\text{s.o}}{=} 1 - P(\liminf_{n\to\infty} A_n^C) \overset{3.4}{=} 1 - \lim_{n\to\infty} P(\bigcap_{k=n}^{\infty} A_k^C)$
$\overset{\text{Unabh.}}{=} 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty}(1 - P(A_k)) \overset{(1-x\leq e^{-x})}{\geq} 1 - \lim_{n\to\infty} \exp\{-\sum_{k=n}^{\infty} P(A_k)\}$
$= 1$, daraus folgt Beh.

(5.15) **Bemerkung**:

i) Analog gilt (mit de Morgan)
  * $\sum_{n=1}^{\infty} P(A_n^C) < \infty \Rightarrow P(\liminf_{n\to\infty} A_n) = 1$.
  * $(A_n)_n$ stoch. unabhängig $\sum_{n=1}^{\infty} P(A_n^C) = \infty \Rightarrow P(\liminf_{n\to\infty} A_n) = 0$.
ii) Für stoch. unabhängige Ereignisse $(A_i)_i$ gilt stets
$P(\limsup_{n\to\infty} A_n)$, $P(\liminf_{n\to\infty} A_n) \in \{0, 1\}$.
iii) Sei $(A_n)_n \subset \mathfrak{A}$, falls eine unabhängige Teilfolge $(A_{n_k})_k$ von $(A_n)_n$ existiert mit $\sum_{k=1}^{\infty} P(A_{n_k}) = \infty$, dann folgt $P(\limsup_{n\to\infty} A_n) = 1$.

(5.16) **Beispiel**:

i) Unendliche Folge von Urnen, je eine Kugel wird gezogen; Urne n enthält eine weiße und $n-1$ schwarze Kugeln,
$A_n \hat{=}$ gezogene Kugel aus der n-ten Urne ist weiß
$\Rightarrow P(A_n) = \frac{1}{n}$ (Ziehung unabhängig)
$\Rightarrow \sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\limsup_{n\to\infty} A_n) = 1$.
Das heißt, mit Wahrscheinlichkeit 1 werden unendlich viele weiße Kugeln gezogen.

ii) Wie i), aber n-te Urne enthält n^2-1 schwarze Kugeln
$\Rightarrow \sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(\limsup_{n\to\infty} A_n) = 0$.
Das heißt, mit Wahrscheinlichkeit 1 werden nur endlich viele weiße Kugeln gezogen.
Beachte: Die Anzahl kann nicht so gewählt werden,
dass $P(\limsup_{n\to\infty} A_n) \in (0,1)$.

(5.17) i) Unabhängiges Werfen eines Würfels.
Gesucht: Wahrscheinlichkeit, unendlich oft die 6 zu würfeln.
$A_n \hat{=}$ 6 im n-ten Wurf, d.h., $P(A_n) = \frac{1}{6}\ \forall n \in \mathbb{N}$
$(A_n)_n$ unabhängig, $\sum P(A_n) = \infty \Rightarrow P(\limsup_{n\to\infty} A_n) = 1$
Beachte: überabzählbares Modell.
zum Beispiel Münzwurf $\Omega = \{(\omega_n)_{n\in\mathbb{N}}; \omega_n \in \{0,1\}, n \in \mathbb{N}\}$
$\rightarrow$ Dualdarstellung der reellen Zahlen in $(0,1)$.

ii) Gesuchte Wahrscheinlichkeit, unendlich oft 2 Sechsen hintereinander zu werfen.
$B_{n,n+1} \hat{=}$ 6 im n-ten Wurf und im $(n+1)$-ten Wurf, nicht unabhängig mit n.
$P(B_{n,n+1}) = \frac{1}{36}$, aber $(B_{2n,2n+1})_n$ unabhängige Familie.
$\Rightarrow$ (5.15) iii) $P(\limsup_{n\to\infty} B_{n,n+1}) = 1$.

(5.18) **Bemerkung** (Produktexperimente):
Idee: Modelle $(\Omega_i, \mathfrak{A}_i, P_i)$ bekannt, $1 \le i \le n$.
zum Beispiel: Ziehen mit Zurücklegen aus der Urne, oder Würfeln.
Ziel: Modell für Experiment, das aus unabhängiger Hintereinanderausführung der Teilexperimente besteht, zum Beispiel n-maliges Ziehen, n-maliger Würfelwurf.
$\Rightarrow \Omega = \{(\omega_1, ..., \omega_n); \omega_i \in \Omega_i, 1 \le i \le n\} = \prod_{j=1}^{n} \Omega_i$
(Ω_i nicht notwendig identisch)$= \times_{i=1}^{n} \Omega_i$.

(5.19) **Definition**:
Für diskrete WR $(\Omega_i, \mathfrak{A}_i, P_i)$, $1 \le i \le n$, heißt $(\Omega, \mathfrak{A}, P)$ mit
$\Omega = \prod_{i=1}^{n} \Omega_i = \{\omega = (\omega_1, ..., \omega_n); \omega_i \in \Omega_i, 1 \le i \le n\}$
und P definiert durch $P(\omega) = \prod_{i=1}^{n} P_i(\omega_i)$
$(P := \times_{i=1}^{n} P_i)$
($\mathfrak{A}$ ist Potenzmenge zu Ω)
Produkt der WR $(\Omega_i, \mathfrak{A}_i, P_i)$, $1 \le i \le n$,
Bezeichnung $(\Omega, \mathfrak{A}, P) = \bigotimes_{i=1}^{n} (\Omega_i, \mathfrak{A}_i, P_i)$.

(5.20) **Beispiel**:

Binomialverteilung, n-fache Wiederholung eines Experiments mit den Ausgängen 0 oder 1.
$\Omega = \{0,1\}^n$, $P(\omega) = p^k \cdot (1-p)^{n-k}$, falls k die Anzahl der Einsen in $\omega = (\omega_1, ..., \omega_n)$ und $P_i(1)$, $1 \leq i \leq n$.
Bezeichnung: p Erfolgswahrscheinlichkeit.
$\omega_i = 1$ Erfolg im i-ten Teilexperiment n.
Bernoulli-Experimente mit Erfolgswahrscheinlichkeit p.
$E_k = \{\omega \in \Omega; \sum_{i=1}^n \omega_i = k\} \hat{=}$ gesamter Erfolg; $k \leq n$.
$P(E_k) \stackrel{\text{s.o.}}{=} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$
Die durch $P(E_k)$ (E_i's sind disjunkt) definierte WV auf $\{0, 1, ..., n\}$ heißt Binomialverteilung.

2 Zufallsvariablen auf diskreten W'räumen

2.6 Zufallsvariablen

Zufallsvorgänge werden beschrieben durch WR $(\Omega, \mathfrak{A}, P)$.
Dabei häufig:
$\omega \in \Omega$ selbst nicht von Interesse, sondern Funktion X von ω.
zum Beispiel:

ω:	n-facher Münzwurf	$\rightarrow$	$X(\omega)$:	Anzahl „Zahl“
ω:	Telefongespräch	$\rightarrow$	$X(\omega)$:	Dauer
ω:	Aktienmarkt	$\rightarrow$	$X(\omega)$:	Kurs einer Aktie

Frage: Ist Grundraum vollständig beschreibbar?
Nicht alle Einflussvariablen beschreibbar?
Also: neue Form der Modellierung $\rightarrow$ Zufallsvariablen.

(6.1) **Beispiel**:

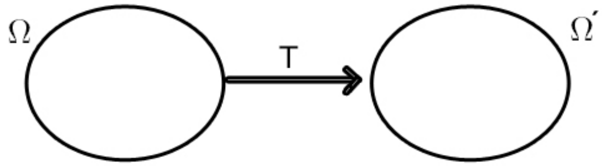
i) Fortsetzung von (5.20): n-facher unabhängiger Münzwurf $\Omega = \{0,1\}^n$
$P(\omega) = p^k \cdot (1-p)^{n-k}$, falls k Einsen in ω.
Betrachte Abbildung X: $\begin{cases} \Omega & \rightarrow & \mathbb{N}_0 \\ \omega & \mapsto & \sum_{i=1}^n \omega_i \end{cases}$
$\Rightarrow P(\{\omega \in \Omega; X(\omega) = k\}) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}, 0 \leq k \leq n.$
Wegen $\sum_{k=0}^n \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = 1$ definiert $P'(k) := \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$ eine WV auf $\Omega' = \{0, 1, ..., n\}$, i.e. (s.o.) Binomialverteilung $b(\omega, p)$.

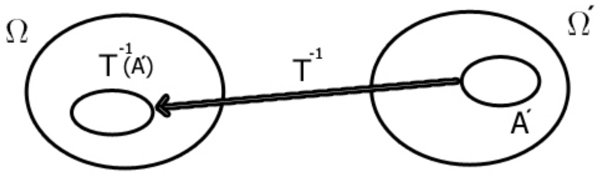
ii) Fortsetzung von (1.7), Problem von de Méré.
$\Omega = \{(\omega_1, \omega_2, \omega_3); \omega_i \in \{1, ..., 6\}, i = 1, 2, 3\}$ und Laplaceverteilung auf Ω, oder $\Omega = \{1, ..., 6\}^3$.
Schon gesehen: $P(\sum = 11) = \frac{27}{216} > \frac{25}{216} = P(\sum = 12)$.
Allgemeiner:
Gesucht ist $q_r := P(\{\omega \in \Omega, \omega_1 + \omega_2 + \omega_3 = r\})$ für $3 \leq r \leq 18$, klar $\sum_{r=3}^{18} q_r = 1$,
d.h., $q_{11} = \frac{27}{216}$, $q_{18} = \frac{25}{216}$.
(weiterhin zum Beispiel: $q_{10} = q_{11}$, $q_9 = q_{12}$, $q_8 = q_{13} = \frac{21}{216}$, ...)
Betrachte: X: $\begin{cases} \Omega & \rightarrow & \mathbb{R} \\ (\omega_1, \omega_2, \omega_3) & \mapsto & \omega_1 + \omega_2 + \omega_3 \end{cases}$.
Dann wird durch X eine WV P^X auf $\Omega' = \{3, ..., 18\}$ erzeugt,
wobei $P^X(r) = q_r$.
(Klar: hier keine Laplaceverteilung.)
Beachte:

* Abbildung $X : \Omega \rightarrow \mathbb{R}$, $\mathbb{R}$ nicht abzählbar, aber Ω abzählbar $\Rightarrow$ Bild von Ω unter X ist abzählbar.
* Möglich $\omega_1 \neq \omega_2$, aber $X(\omega_1) = X(\omega_2) = x$.
  Dann $P^X(x) = P(\{\omega \in \Omega; X(\omega) = x\})$.

Struktur in den Beispielen:
Gegeben diskreter WR, Abb. $X : \Omega \rightarrow \mathbb{R}$.

Problem: Ist $P^X(k) := P(\{\omega \in \Omega; X(\omega) = k\})$; $k \in X(\Omega)$.
Oder allgemeiner ist $P^X(B) := P(\{\omega \in \Omega, X(\omega) \in B\}) = P(X^{-1}(B))$,
$B \in \mathcal{P}(X(\Omega))$ ein diskreter WR über $X(\Omega)$ ($\neq \emptyset$, abzählbar)?
Dazu Urbildfunktion betrachten.





(6.2) **Definition**:
Sei $T : \Omega \to \Omega'$ eine Abbildung.
Die Abbildung $T^{-1} : \begin{cases} \mathcal{P}(\Omega') & \to \quad \mathcal{P}(\Omega) \\ \quad A' & \mapsto \quad T^{-1}(A') := \{\omega \in \Omega; T(\omega) \in A'\} \end{cases}$
heißt die zu T gehörige Urbildfunktion.

(6.3) Eigenschaften der in (6.2) definierten Urbildfunktion. Seien $A', B', A_i' \in \mathcal{P}(\Omega')$.

i) $T^{-1}(\emptyset) = \emptyset$, $T^{-1}(\Omega') = \Omega$

ii) $T^{-1}(A' \backslash B') = T^{-1}(A') \backslash T^{-1}(B')$
speziell: $T^{-1}(B'^C) = (T^{-1}(B'))^C$

iii) $T^{-1}(\bigcap_{i \in I} A_i) = \bigcap_{i \in I} T^{-1}(A_i)$

iv) $T^{-1}(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} T^{-1}(A_i')$
speziell für disjunkte B_i'
(Symbol $\sum$ für disj. Vereinigung)
$T^{-1}(\underbrace{\sum_{i \in I} B_i'}_{\dot{\bigcup}_{i \in I} B_i'}) = \sum_{i \in I} T^{-1}(B_i')$.

v) $A' \subset B' \Rightarrow T^{-1}(A') \subset T^{-1}(B')$

vi) Ist $S : \Omega' \to \Omega''$ beliebige Abbildung, dann gilt $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$.

(6.4) **Lemma:**
Sei $(\Omega, \mathfrak{A}, P)$ ein diskreter WR und $X : \Omega \to \mathbb{R}$ eine Abbildung.
Dann ist $P^X : \mathcal{P}(\mathbb{R}) \to [0,1]$ mit $P^X(A) = P(X^{-1}(A)) = P(\{\omega \in \Omega; X(\omega) \in A\})$ eine diskrete WV über (einer Teilmenge von) $\mathbb{R}$, bzw. über $X(\Omega)$.
Beweis:
P^X ist definiert über Urbildfunktion; damit (siehe (6.3)):
$P^X(A) \geq 0 \ \forall A \in \mathbb{R}$,
$P^X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\{\omega \in \Omega; X(\omega) \in \mathbb{R}\}) = P(\Omega) = 1$.

zur σ-Additivität:
Sei $(B_i)_{i\in\mathbb{N}}$ eine Folge disjunkter Teilmengen von $\mathbb{R}$. Dann gilt:

$$P^X(\sum_{i=1}^{\infty} B_i) = P(X^{-1}(\sum_{i=1}^{\infty} B_i)) = P(\sum_{i=1}^{\infty} X^{-1}(B_i)) = \sum_{i=1}^{\infty} P(X^{-1}(B_i)) = \sum_{i=1}^{\infty} P^X(B_i).$$

Für nicht-abzählbare Ω ist Einschränkung an Abbildung X erforderlich, um eine Aussage wie in (6.4) zu treffen.

(6.5) **Definition**:
Eine Abbildung $X : \Omega \to \Omega'$ von einem Messraum $(\Omega, \mathfrak{A})$ in einen anderen Messraum $(\Omega', \mathfrak{A}')$ heißt messbar, falls für alle $A' \in \mathfrak{A}'$ gilt: $X^{-1}(A') \in \mathfrak{A}$.
Bemerkung:
Die Elemente einer σ-Algebra heißen auch messbare Mengen. Dann (6.5) heißt X messbar, falls die Urbilder messbarer Mengen wieder messbar sind.

(6.6) **Definition**:
Eine messbare Funktion im Sinne von (6.5) von einem WR in einen anderen heißt Zufallsvariable (ZV), bzw. Zufallsvektor, falls $\Omega' = \mathbb{R}^n$ (auch Zufallsgröße).
Der Betriff der ZV'en ist zur Modellierung einer der Wesentlichen in der Wahrscheinlichkeitstheorie.
Diese Abbildung beinhaltet das jeweils interessierende Merkmal der jeweiligen Teilaspekte eines eventuell komplexen Modells (vergleiche Einleitung).
Falls Teilapsekt eines Zufallsexperiments durch ZV beschreibar, betrachte nur noch diese ZV und deren Verteilung ohne Rückgriff auf die explizite Gestalt des zugrundeliegenden WR's.

(6.7) **Bemerkung**:

i) Ist $\mathfrak{A} = \mathcal{P}(\Omega)$ wie beim diskreten WR, so ist jede Abbildung $X : \Omega \to \Omega'$ messbar.
ii) Kurzschreibweise:
$\{x \in A\} := X^{-1}(A) = \{\omega \in \Omega; X(\omega) \in A\}$
dann auch kurz: $P(x \in A)$.
iii) Die Komposition messbarer Funktionen ist messbar.

(6.8) **Definition**:
Das Wahrscheinlichkeitsmaß P^X, definiert durch $P^X(A) = P(X^{-1}(A))$, $A \subseteq \mathbb{R}$, heißt Verteilung von X unter P.
Bezeichnung:
X hat Verteilung P^X, X ist verteilt wie P^X, $X \sim P^X$, $X \sim P$.

(6.9) (Beispiel) (binary search):
$\Omega = \{0, 1, ..., 2^n - 1\}$ mit Laplaceverteilung.
(geordnetes Feld der Länge $2^n - 1$ plus zusätzliche Möglichkeiten)
$\omega \geq 1$, $\omega \in \Omega$: mögliche Platznummer des Schlüsselelements (ist zu suchen).

$\omega = 0$: gesuchtes Element ist nicht im Feld vorhanden.

$\bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \underbrace{\bullet}_{\substack{\uparrow \\ \text{Start}}} \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet$ hier für $n = 3$, also $2^3 - 1$ Schritte.

$A_1 \hat{=}$ Schlüsselelement wird in genau einem Schritt gefunden.
$= \{2^{n-1}\}$ (Element in der Mitte des Feldes)
$A_2 \hat{=}$ Schlüsselelement wird in genau zwei Schritten gefunden.
$= \{2^{n-2}, 3 \cdot 2^{n-2}\}$
$A_k \hat{=}$ Schlüsselelement wird in genau k Schritten gefunden.
$= \{(2j-1)2^{n-k}; 1 \leq j \leq 2^{k-1}\}, 1 \leq k \leq n$.
Damit:
$|A_k| = 2^{k-1}$ und $P(A_k) = \frac{2^{k-1}}{2^n}$, $1 \leq k \leq n$.
...
Von der Laplaceverteilung über $0, 1, ..., 2^n - 1$,
das heißt, P(Element nicht im Feld)$= \frac{1}{2^n}$ (als implizite Vor.).
Weiter:
$B_k \hat{=}$ Suche nach höchstens $k \leq n$ Schritten beendet.
$B_k = \sum_{i=1}^{k} A_i$, also

$$P(B_k) = \sum_{i=1}^{k} P(A_i) = \frac{1}{2^n} \cdot \sum_{i=1}^{k} 2^{i-1} = \frac{2^k - 1}{2^n}.$$

(ungünstigster Fall: B_n; $P(< n \text{ Schritte}) = P(B_{n-1}) = \frac{2^n - 1}{2^n} < \frac{1}{2}$.
Bis hier alte Beschreibung.
Nun:
Sei ZV X definiert durch $X(\omega) = \begin{cases} k & \omega \in A_k, \\ n & \omega \in A_0 \cup A_n [A_0 = \{0\}]. \end{cases}$
X zählt die Schritte bis zum Abbruch des Verfahrens und ordnet jeder Platznummer ω, die in genau $k \in \{1, ..., n-1\}$ Schritten erreichbar ist, den Wert k zu.
$X = n$, falls Schlüsselelement nicht in Liste (d.h., alles abgesucht), bzw. maximale Schrittzahl benötigt wird.
$X(\omega) = \begin{cases} k & \omega \in A_k, \\ n & \omega \in A_0 \cup \{A_n\}. \end{cases}$
Verteilung P^X ist diskret und bestimmt durch
$P(X = k) = \begin{cases} P(A_k) = 2^{k-1-n} & 1 \leq k < n, \\ P(A_0 \cup A_n) = P(A_0) + P(A_n) = \frac{1}{2}n + \frac{1}{2} & k = n. \end{cases}$
Entsprechend (vgl. Beispiel):
$P(X \leq k) = \sum_{i=1}^{k} P(X = 1) = \frac{2^k - 1}{2^n}$.

(Beispiel, Fortsetzung von (6.1)):

i) ZV X ist Binomialverteilung, falls
$P^X(k) = P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$.

ii) ZV X ist Poissonverteilung, falls für ein $\lambda > 0$ gilt:
$P^X(k) = P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$, $k \in \mathbb{N}_0$.

iii) Seien $\Omega \neq 0$, P diskrete WV über Ω, $A \subset \Omega$.
Die Funktion $I_A : \Omega \to \mathbb{R}$ (statt I_A auch $1|_A$),
definiert durch $I_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \notin A, \end{cases}$
heißt Indikatorfunktion von A und ist ZV.
Dabei $I_A \sim b(1,p)$ mit $p = P(A)$, denn
$P(I_A = 0) = P(\{\omega \in \Omega; I_A(\omega) = 0\}) = P(A^C) = 1 - p$,
$P(I_A = 1) = P(\{\omega \in \Omega; I_A(\omega) = 1\}) = P(A) = p$.
Weiterhin gilt:
$I_{A\cup B} = \max(I_A, I_B)$,
$I_{A\cup B} = I_A + I_B$, falls $A \cap B = \emptyset$,
$I_{A\cap B} = \min(I_A, I_B) = I_A \cdot I_B$,
$I_{A^C} = 1 - I_A$.
Zum Zusammenhang ZV'en $X_1, ..., X_n$ und ZVektor $(X_1, ..., X_n)$:

- Aufbau eines Vektors durch verschiedene ZV'en.
- Zerlegung eines Vektors in Komponenten.

(6.10) (Beispiel, verallgemeinerte Bernoulli-Experimente):
Zufallsexperiment liefert eines von $m \geq 2$ möglichen Ergebnissen A_i
(zum Beispiel Maschine fällt auf Grund von Defekt i aus, $1 \leq i \leq n$).
Voraussetzung:
$A_1, ..., A_m$ disjunkt, $P(A_j) = p_j$, $1 \leq j \leq m$, $\sum_{j=1}^m p_j = 1$.
Nun: n-malige Versuchswiederholung $\to \Omega = \{1, ..., m\}^n$.
Frage: $P(\underbrace{|A_1| = k_1, ..., |A_m| = k_m}_{\text{Verteilung des Defekts}}, \sum_{j=1}^m k_j = n)$.
Setze X_j (ZV), zählt die Defekte A_j bei m Versuchen, $1 \leq j \leq m$.
Also $P(X_1 = k_1, ..., X_m = k_m)$
$= P(\{\omega \in \Omega; X_1(\omega) = k_1, ..., X_m(\omega) = k_m\})$
$= P^{X_1,...,X_m}(\{(k_1, ..., k_m)\})$
$= \binom{n}{k_1} \cdot \binom{n-k_1}{k_2} \cdot ... \cdot \binom{n-k_1-...-k_{m-1}}{k_m} \cdot p_1^{k_1} \cdot ... \cdot p_m^{k_m}$
$= \frac{n!}{k_1! \cdot ... \cdot k_m!} \cdot \prod_{i=1}^m p_i^{k_i}, k_j \in \mathbb{N}_0, 1 \leq j \leq m, \sum_{j=1}^m k_j = n$.
Durch diesen Ausdruck ist eine endliche diskrete WV über
$\{(X_1, ..., X_m) \in \mathbb{R}^m; X_j \in \mathbb{N}_0, j \in \{1, ..., m\}, \sum_{j=1}^m X_j = n\}$,
kurz über $\mathbb{R}^m$ erklärt.

Bezeichnung:
$(X_1, ..., X_m)$ genügt der Multinomial- / Polynomialverteilung mit den Paramtern $n, p_1, ..., p_m$.
Bez.: $M(n, p_1, ..., p_m)$.
Speziell: $m = 2$: $M(n, p_1, p_2) = b(m, p_1)$, da $p_1 + p_2 = 1$.

(6.11) **Definition**:
Seien $X_1, ..., X_m$ ZV'en. Die Verteilung von $X = (X_1, ..., X_m)$ heißt die *gemeinsame Verteilung* der ZV'en $X_1, ..., X_m$.

Bezeichnung:
$P^X = P^{(X_1,...,X_m)}$.
Die Verteilung von $(X_{i_1},...,X_{i_l}), 1 \leq i_1 < ... < i_l \leq m, l < m$, heißt l-dimensionale Randverteilung (Marginalverteilung) zu $(i_1,...,i_l)$. Die Verteilung von X_i heißt die i-te Randverteilung (Marginalverteilung), $1 \leq i \leq m$.
Bemerkung:
Die gemeinsame Verteilung ist durch die Angabe aller Wahrscheinlichkeiten $P(X_1 \in A_1, ..., X_m \in A_m)$, $A_i \in \mathfrak{A}_i$ bestimmt.

(6.12) **Lemma**:
P^X sei eine diskrete WV über $\mathbb{R}^n$.
Randverteilung von $(X_{i_1},...,X_{i_m})$ wird bestimmt durch
$P^{(X_{i_1},...,X_{i_m})}(B) = P((X_{i_1},...,X_{i_m})^{-1}(B)) = P((X_1,...,X_n)^{-1}(B \times \mathbb{R}^{n-m}))$
mit $B \in \mathcal{P}(\mathbb{R}^m)$
und $B \times \mathbb{R}^{n-m} := \{(X_1,...,X_m) \in \mathbb{R}^n; (X_{i_1},...,X_{i_m}) \in B\}$
zum Beispiel $n = 2$:
$P^{X_1}(B) = P^{X_1,X_2}(B \times \mathbb{R})$
Also $P^{(X_{i_1},...,X_{i_m})}(B) = P^X(B \times \mathbb{R}^{n-m})$.

(6.13) (Beispiel:) (Fortsetzung von Beispiel (6.10))
Bestimmung der ersten Randverteilung; sei $B = \{k\}$.

$$P^{X_1}(B) = P^X(B \times \mathbb{R}^{m-1}) = \sum_{k_2} ... \sum_{k_m} \frac{n!}{k_2! \cdot ... \cdot k_m!} \cdot \prod_{i=2}^{m} p_i^{k_i} \cdot \frac{p_1^k}{k!}$$

$$\left[\sum_{j=2}^{m} k_j = n-k\right]$$

$$= \frac{n!}{(n-k)!} \cdot \sum_{k_2} ... \sum_{k_m} \frac{(n-k)!}{k_2! \cdot ... \cdot k_m!} \cdot \prod_{i=2}^{m} \left(\frac{p_i}{1-p_1}\right)^{k_i} \cdot (1-p_1)^{k_i} \cdot \frac{p_1^k}{k!}$$

$$= \binom{n}{k} \cdot p_1^k \cdot \underbrace{\sum_{k_2} ... \sum_{k_m} \frac{(n-k)!}{k_2! \cdot ... \cdot k_m!} \cdot \prod_{i=2}^{m} \left(\frac{p_i}{1-p_1}\right)^{k_i}}_{=1,(\text{Multinormalverteilung mit } (n-k)\cdot\left(\frac{p_2}{1-p_1}\right)\cdot...\cdot\left(\frac{p_m}{1-p_1}\right))} \cdot \prod_{j=2}^{m}(1-p_1)^{k_j}$$

$$= \binom{n}{k} \cdot p_1^k \cdot (1-p_1)^{n-k}.$$

(6.14) **Bemerkung und Beispiel**:
Die eindimensionale Randverteilung legen die gemeinsame Verteilung nicht eindeutig fest.
Seien X, Y ZV'en auf WR $(\Omega, \mathfrak{A}, P)$ mit
$X(\Omega) = Y(\Omega) = \{0,1\}^2$
$[(0,0),(0,1),(1,0),(1,1)$ Werte für $X, Y]$

$X, Y : \Omega \to \mathbb{R}^2, X = (X_1, X_2), Y = (Y_1, Y_2)$ mit
$P^X(0,1) = P^X(1,0) = \frac{1}{2}$
$P^X(0,0) = P^X(1,1) = 0$
$P^Y(0,0) = P^Y(1,1) = \frac{1}{2}$
$P^Y(0,1) = P^Y(1,0) = 0.$

$X_2 \backslash X_1$	0	1
0	0	$\frac{1}{2}$
1	$\frac{1}{2}$	0

$Y_2 \backslash Y_1$	0	1
0	$\frac{1}{2}$	0
1	0	$\frac{1}{2}$

$\Rightarrow P^X$ und P^Y sind verschiedene 2-dimensionale Verteilungen,
aber $P^{X_1} = P^{Y_1}$ und $P^{X_2} = P^{Y_2}$.

$$P^{X_1}(j) = P^X(\{j\} \times \Omega) = P^X(j,0) + P^X(j,1) = \begin{cases} 0 + \frac{1}{2} & j = 0 \\ \frac{1}{2} + 0 & j = 1 \end{cases}$$

$$P^{Y_1}(j) = P^Y(\{j\} \times \Omega) = P^Y(j,0) + P^Y(j,1) = \begin{cases} \frac{1}{2} + 0 & j = 0 \\ 0 + \frac{1}{2} & j = 1 \end{cases}$$

$P^{X_2} = P^{Y_2}$ analog.

Nun unabhängig von ZV'en.

(6.15) **Definition**:
Eine Familie von ZV'en $X_i : (\Omega, \mathfrak{A}, P) \to (\Omega_i, \mathfrak{A}_i, P^{X_i}), i \in I$ heißt *stochastisch unabhänging* (oder die ZV'en $X_i, i \in I$ heißen stochastisch unabhängig), falls die Mengensysteme $X_i^{-1}(\mathfrak{A}_i)$ unabhängig sind, d.h., jedes Repräsentantensystem $B_i \in X_i^{-1}(\mathfrak{A}_i), i \in I$, bildet eine unabhängige Familie von Ereignissen.

(6.16) Die ZV'en $X_i, i \in I$ sind stochastisch unabhängig
$\Leftrightarrow P(\bigcap_{i=j}\{X_i \in A_i\}) = \prod_{i=j} P(X_i \in A_i) \; \forall J \subseteq I, |J| < \infty$ und $\forall A_i \in \mathfrak{A}_i, i \in J$.

(6.17) **Satz**:
Sind die ZV'en $X_i : \Omega \to \Omega_i, i \in I$ stochastisch unabhängig und sind die Abbildungen $f_i : \Omega_i \to \Omega_i'$ messbar, so sind die ZV'en $f_i \circ X_i, i \in I$ stochastisch unabhängig.
Weiterhin:
Seien $I_j \subset I$ für $j \in J$ disjunkte Teilmengen und $g_i : \times_{i \in I_j} \Omega_i \to \Omega_j^*$ messbar.
$\Rightarrow g_j \circ (X_i, i \in I_j), j \in J$, stochastisch unabhängig (messbare Funktionen von ZV'en mit disjunkten Indexmengen).

(Beispiel:)
X_1, X_2, X_3 unabhängig $\Rightarrow$
$X_2, (X_1, X_3)$ unabhängig
$X_2^2, (X_1 - X_3)$ unabhängig

(6.18) **Lemma**:
Sei $(\Omega, \mathfrak{A}, P)$ ein diskreter WR. Dann gilt:
$X_i, i \in I$, stoch. unabh.

$$\Leftrightarrow P(X_j = x_j, j \in J) = \prod_{j \in J} P(X_j = x_j) \; \forall x_j \in X_j(\Omega), \; \forall j \in J, \; \forall J \subseteq I, |J| < \infty.$$

(6.19) **Bemerkung**:
Es gilt mit (6.18):
$P^X = P^{(X_1,...,X_n)} = \times_{i=1}^n P^{X_i}$, also:
$X_1, ..., X_n$ stoch. unabhängig.

$$\Leftrightarrow P^{(X_1,...,X_n)}(\times_{i=1}^n A_i) = \prod_{i=1}^n P(X_i \in A_i) = \prod_{i=1}^n P^{X_i}(A_i) \ \forall A_i \in \mathfrak{A}_i$$

$$\left[P^X(\times_{i=1}^n A_i) = P^X(\{\omega = (\omega_1, ..., \omega_n); \omega_i \in A_i\}) = \sum_{\omega;\omega_i \in A_i} \prod_{i=1}^n P^{X_i}(\omega_i) \overset{\text{(s.o.)}}{=} \prod_{i=1}^n P^{X_i}(A_i) \right].$$

(6.20) **Bemerkung**:
Es gilt:
Die ZV'en X und Y sind stoch. unabhängig unter P.
$\Rightarrow$ ZV'en $f(X)$ und $g(Y)$ sind stoch. unabh. $\forall f, g : \mathbb{R}^1 \to \mathbb{R}^1$, f, g messbar.
Jetzt: Verteilung der Summe zweier unabh. ZV'en.

(6.21) **Satz**:
X, Y seien stoch. unabh. ZV'en auf $\mathbb{Z}$ mit den Zähldichten f, bzw. g.
(d.h., $P(X = n) = f(n)$, $P(Y = m) = g(m)$)
Dann hat $X + Y$ die Zähldichte h gegeben durch

$$h(k) = \sum_{j \in \mathbb{Z}} f(j)g(k-j) = \sum_{j \in \mathbb{Z}} f(k-j)g(j), k \in \mathbb{Z} = P(X + Y = k).$$

Bezeichnung:
h ist Faltung der Dichten f und g: $h = f * g$.
Beweis:
Es ist $\sum_{j \in \mathbb{Z}} P(X = j) = \sum_{j \in \mathbb{Z}} P(Y = j) = 1$.

$$\Rightarrow P(X + Y = k) \underset{\text{lokale W'}}{=} \sum_{j \in \mathbb{Z}} P(X + Y = k, Y = j) = \sum_{j \in \mathbb{Z}} P(X + j = k, Y = j)$$

$$\underset{\text{stoch. unabh.}}{=} \sum_{j \in \mathbb{Z}} \underbrace{P(X = k - j)}_{f(k-j)} \cdot \underbrace{P(Y = j)}_{g(j)}.$$

(6.22) (Beispiel:)

(i) Seien X, Y unabhängige $b(1, p)$-vert. ZV'en,
d.h., $P(X = 0) = 1 - p$, $P(X = 1) = p$.

$$\Rightarrow P(X + Y = k) = \begin{cases} P(X = 0) \cdot P(Y = 0) & k = 0 \\ P(X = 0) \cdot P(Y = 1) + P(X = 1) \cdot P(Y = 0) & k = 1 \\ P(X = 1) \cdot P(Y = 1) & k = 2 \end{cases}$$

$$= \begin{cases} (1-p)^2 & k=0 \\ 2p(1-p) & k=1 \\ p^2 & k=2 \end{cases} = \binom{2}{k} p^k (1-p)^{2-k}, k \in \{0,1,2\}$$

i.e. Binomialverteilung $b(2, p)$ per Induktion.
Seien $X_1, ..., X_n$ stoch. unabhängig $b(1-p)$-vert. ZV'en
$\Rightarrow \sum_{i=1}^n X_i$ besitzt eine $b(n,p)$-Verteilung.

(ii) Seien X, Y unabh. $X \sim_{\text{po}} (\lambda)$, $Y \sim_{\text{po}} (\mu)$, $\lambda, \mu > 0$ (siehe (2.18)).

$$P(X+Y=k) = \sum_{j \in \mathbb{N}_0} P(X=k-j)P(Y=j) = \sum_{j=0}^{k} \frac{\lambda^{k-j}}{(k-j)!} \cdot e^{-\lambda} \cdot \frac{\mu^j}{j!} \cdot e^{-\mu}$$

$$= \frac{e^{-(\lambda+\mu)}}{k!} \cdot \underbrace{\sum_{j=0}^{k} \binom{k}{j} \lambda^{k-j} \mu^j}_{(\lambda+\mu)^k}$$

d.h., $X + Y \sim_{\text{po}} (\lambda + \mu)$.
Bisher zur Beschreibung des Zufalls
Zähldichte $p : \sum_{\omega \in \mathbb{Z}} p(\omega) = 1$
WV $P : \mathfrak{A} \to [0,1]$, $P(A) = \sum_{\omega \in A} p(\omega), A \in \mathfrak{A}$.

Weiteres Hilfsmittel:

(6.23) **Definition**:
Seien $(\mathfrak{A}^0, \mathfrak{A}, P)$ ein WR und X eine ZV mit WV P^X.
Die Funktion $F^X : \begin{cases} \mathbb{R}^1 & \to & \mathbb{R}^1 \\ x & \mapsto & P^X((-\infty, x]) \end{cases}$
heißt die zu P^X gehörige Verteilungsfunktion von X.
(Wenn Zusammenhang klar: F Verteilungsfunktion von X,
X verteilt nach $F, ..., X \sim F$).

(6.24) **Bemerkung**:

(i) Sei P^X die Zähldichte von P^X, dann ist

$$F^X(x) = P^X((-\infty, x]) = \sum_{\omega \leq x} P^X(\omega) = \sum_{\substack{\omega \leq x \\ \omega \in \operatorname{supp} P^X}} P^X(\omega), X \in \mathbb{R}.$$

(ii) Für $P^X((-\infty, x])$ schreibt man (siehe (6.7)):
$P^X((-\infty, x]) = P(X^{-1}((-\infty, x]))$
$= P(\{\omega \in \Omega; X(\omega) \in (-\infty, x]\}) = P(X \leq x)$,
d.h., $F^X(x)$ ist die Wahrscheinlichkeit dafür, dass die ZV X Werte $\leq x$ annimmt.

(6.25) **Lemma**:
Sei F^X die zu P^X gehörige Verteilungsfunktion. Dann gilt:

a) (i) F^X ist monoton wachsend.
(ii) F^X ist rechtsseitig stetig.
(iii) $\lim_{x\to\infty} F^X(x) = 1$, $\lim_{x\to-\infty} F^X(x) = 0$.

b) P^X ist durch F^X eindeutig bestimmt.

Beweis (nur a)):

(i) Seien $x_1 \le x_2$ beliebig. Wegen $(-\infty, x_1] \subset (-\infty, x_2]$
ist $F^X(x_1) = P^X((-\infty, x_1]) \le P^X((-\infty, x_2]) = F^X(x_2)$.

(ii) Sei $(x_n)_{n\in\mathbb{N}}$ eine Folge mit $x_n \downarrow x$, $n \to \infty$ und $(x_n)_n$ mon. fallend.
Mit $A_n := (-\infty, x_n]$, $n \in \mathbb{N}$ gilt dann:
$A_n \supset A_{n+1}$, $n \in \mathbb{N}$ und $\bigcap_{n=1}^\infty A_n = (-\infty, x] =: A$.

$$\underset{\text{Stetigkeit von oben}}{\Rightarrow} \lim_{n\to\infty} F^X(x_n) = \lim_{n\to\infty} P^X(A_n) = P^X(A) = F^X(x).$$

(iii) Sei $(x_n)_{n\in\mathbb{N}}$ monoton wachsend $x_n \to \infty$, $n \to \infty$.
Dann: $A_n \uparrow \mathbb{R}$ für $A_n := (-\infty, x_n]$.

$$\Rightarrow \lim_{n\to\infty} F^X(x_n) = \lim_{n\to\infty} P^X(A_n) = P^X(\mathbb{R}) = 1.$$

Zweiter Grenzwert analog.

(6.26) **Bemerkung**:

(i) F^X ist durch P^X ebenfalls eindeutig bestimmt, denn:
F^X Verteilungsfunktion zu P' mit Zähldichte p'.
$\Rightarrow p'(x) = F^X(x) - F^X(x-) = p(x)\ \forall x \in \mathbb{R}$.

(ii) Es gilt $p^X(x) = F^X(X) - F^X(x-)$ (siehe oben),
d.h., F^X stetig in $x \in \mathbb{R} \Leftrightarrow p^X(x) = 0$ (sonst Sprung)
($p^X(x) = 0 \Rightarrow F^X$ linksseitig stetig; F^X rechtsseitig stetig
$\Rightarrow F^X$ stetig.)
Weiter:
$\exists$ höchstens abzählbar viele Punkte $x \in \mathbb{R}$ mit $p^X(x) > 0$ (sonst $\sum p(\omega) = \infty$).
$\Rightarrow \exists$ höchstens abzählbar viele Unstetigkeitsstellen von F^X.

(iii) zu (6.25)(b): WV über $\mathbb{R}$ ist eindeutig bestimmt durch:
$P^X((-\infty, x]), x \in \mathbb{R}$.

(iv) Sei $\{x_1, x_2, ...\}$, $x_i < x_{i+1}$, $i \in \mathbb{N}$ eine Abzählung des Trägers von P^X. Dann:

$$P^X((-\infty, x_{i+1})) = \underbrace{P^X((-\infty, x_i])}_{F^X(x_i)} + \underbrace{P^X((x_i, x_{i+1}))}_{=0,\text{ da nicht in Träger}},$$

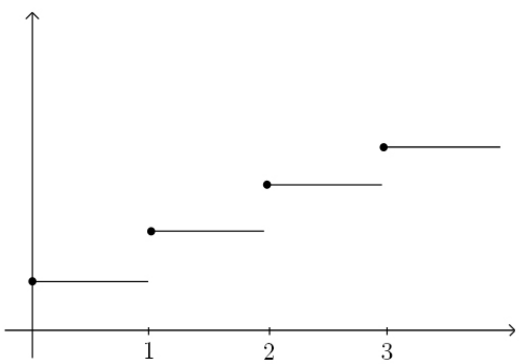
d.h., F^X ist eine „Treppenfunktion“:
Sprünge an den Trägerpunkten, konstant sonst.

(6.27) (Beispiel:)
Verteilungsfunktion der $b(5, \frac{1}{2})$-Verteilung.

$$p^X(x) = \begin{cases} \binom{5}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x} & x \in \{0, ..., 5\}, \\ 0 & \text{sonst.} \end{cases}$$

D.h., $\operatorname{supp}(p^X) = \{0, ..., 5\}$

$$F^X(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{2^5} \cdot \sum_{i=0}^{\lfloor x \rfloor} \binom{5}{i} & 0 \leq x \leq 5, \\ 1 & x > 5. \end{cases}$$



(6.28) **Definition**:
Sei $X = (X_1, ..., X_n)$ ein Zufallsvektor mit der WV P^X.
Die durch $F^X(x) := P^X((-\infty, x_1] \times ... \times (-\infty, x_n])$, $(x_1, ..., x_n) \in \mathbb{R}^n$ definierte Funktion heißt (multivariante) Verteilungsfunktion.
(n=1: siehe (6.23))

(6.29) **Bemerkung**:
Sind X_1, X_2 stochastisch unabhängig, so gilt:
$F^{(X_1, X_2)}(x_1, x_2) = F^{X_1}(x_1) \cdot F^{X_2}(x_2)$, $(x_1, x_2) \in \mathbb{R}^2$.
Beweis direkt aus Definition mit $A_i = (-\infty, x_i]$.
Insbesondere:
Bei stochastisch unabhängigen ZV'en ist die gemeinsame Verteilungsfunktion eindeutig durch die Verteilungsfunktion der eindimensionalen Randwerte bestimmt.

2.7 Erwartungswerte

(Beispiel:)
Würfelspiel:
Ergebnis $i \in \{1, ..., 6\} \Rightarrow$ Auszahlung i Euro.
Durchschnittlich zu erwartende Auszahlung:
$\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + ... + \frac{1}{6} \cdot 6 = 3,5.$
Bezeichnung:
3,5 ist der Erwartungswert der ZV

$$X : \begin{cases} \{1, ..., 6\} & \rightarrow & \mathbb{R}, \\ i & \mapsto & i, \end{cases}$$

wobei P^X Laplace-Verteilung über $\{1, ..., 6\}$.

(7.1) **Definition**:
Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter WR.

(i) Sei $X(\Omega) \subset \mathbb{R}^+$ oder $X(\Omega) \subset \mathbb{R}^-$.
$EX := E(X) := \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$
heißt Erwartungswert von X (unter P).

(ii) Sei X eine reelle ZV mit
$E(\max(X, 0)) < \infty$ oder $E(\min(X, 0)) > -\infty$.
Dann heißt $EX := \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$
der Erwartungswert von X (unter P).

(7.2) **Bemerkung**:

(i) Für nichtnegative ZV ist Erwartungswert immer wohldefiniert, $EX = \infty$ erlaubt.

(ii) Für ZV mit postitiven und negativen Werten muss Wohldefinition der Reihe $\sum_\omega X(\omega) P(\{\omega\})$ gewährleistet werden.
Werden nur endliche Erwartungswerte betrachtet, kann absolute Konvergenz $(\sum_\omega |X(\omega)| P(\omega) < \infty)$ gefordert werden.
($\rightarrow$ Änderung der Summationsreihenfolge erlaubt!)

(iii) Im Folgenden wird stets die Wahrscheinlichkeit der auftretenden Erwartungswerte vorausgesetzt.

(iv) EX hängt nur von der Verteilung von X ab:
Sei $x_1, x_2, ...$ eine Abzählung von $X(\Omega)$, dann

$$EX = \sum_{i=1}^{\infty} \sum_{\{\omega; X(\omega) = x_i\}} X(\omega) \cdot P(\{\omega\}) = \sum_{i=1}^{\infty} x_i \cdot P(X = x_i) = \sum_{i=1}^{\infty} x_i \cdot P^X(x_i).$$

D.h., EX kann ebenso über $\sum x_i P^X(x_i)$ erklärt werden:
$EX = \sum_{x \in \mathbb{R}} P^X(x)$, $(P^X(x) = 0$, falls $x \notin \text{supp}(P))$.

(v) Erwartungswerte (als mögliche Kenngrößen) dienen dem Vergleich von Verteilungen (an Hand einige Kenngrößen).

(7.3) (Beispiele:)

(i) Sei $X \sim b(n,p)$:

$$EX = \sum_{k=0}^{n} k \cdot P^X(k) = \sum_{k=1}^{n} \underbrace{k \cdot \binom{n}{k}}_{n \cdot \binom{n-1}{k-1}} p^k (1-p)^{n-k}$$

$$= n \cdot p \cdot \sum_{k=1}^{n} \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{n-k}$$

$$n \cdot p \cdot \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k \cdot (1-1)^{(n-1)-k}}_{=1,\text{ denn } b(n-1,p)\text{-Vert.}} = n \cdot p.$$

(ii) Sei $X \sim_{\text{po}} (\lambda)$.

$$EX = \sum_{k=0}^{\infty} k \cdot p^X(k) = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

$$= \lambda \cdot \underbrace{e^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{1} = \lambda.$$

(iii) Fortsetzung von (6.9) „binary search“
ZV X zählt Schritte bis zum Abbruch des Verfahrens $X(\Omega) \in \{1, ..., n\}$.
Dort gesehen:

$$P(X = k) = \begin{cases} 2^{k-1-n} & 1 \leq k \leq n-1, \\ \frac{1}{2^n} + \frac{1}{2} & k = n. \end{cases}$$

$$\Rightarrow EX = \sum_{i=1}^{n} iP(X = i) = \sum_{i=1}^{n-1} i \cdot 2^{i-1-n} + n \cdot \left(\frac{1}{2^n} + \frac{1}{2}\right)$$

$$= \sum_{i=1}^{n} i \cdot 2^{i-1-n} + \frac{n}{2^n} = ... = n - 1 + \frac{n+1}{2^n},$$

d.h., die erwartete Schrittzahl bis zum Abbruch des Algorithmus ist - für große n - praktisch nur um Einen besser als im worst case.
($\rightarrow$ average case Analyse)

(iv) ZV X mit Verteilung P^X
Träger: $\mathbb{N}$; $P^X(n) =$const$\cdot \frac{1}{n^2}$
$\left(\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}\right.$, also const $= \left.\frac{6}{\pi^2}\right)$
$\Rightarrow EX = \sum_{n=1}^{\infty} n \cdot \text{const} \cdot \frac{1}{n^2} = \infty$.

(v) ZV X mit Verteilung P^X.
Träger: $\{x_i, i \in \mathbb{N}; x_i = (-1)^i \cdot \frac{2^i}{i}\}$
und $P^X(x_i) = \frac{1}{2^i}$, $(\sum_{i=1}^{\infty} \frac{1}{2^i} = 1)$.
Dann:
$E(\max(X,0)) = \sum_{i=1}^{\infty} \frac{2^{2i}}{2i} \cdot \frac{1}{2^{2i}} = \infty$,
$E(\min(X,0)) = -\sum_{i=1}^{\infty} \frac{2^{2i-1}}{2i-1} \cdot \frac{1}{2^{2i-1}} = -\infty$,
d.h., Erwartungswert von X unter P existiert nicht.
Aber:
$\sum_{i=1}^{\infty} x_i P^X(x_i) = \sum_{i=1}^{\infty} (-1)^i \cdot \frac{1}{i} = -ln(2)$
hängt von der Summationsreihenfolge ab (nicht absolut konvergent).

(vi) Sei $X = I_A$. Dann

$$EI_A = \sum_{\omega \in \Omega} I_A(\omega) \cdot P(\omega) = \sum_{\omega \in A} P(\omega) = P(A).$$

Für ZV'en mit Werten in $\mathbb{N}_0$ gibt es andere Berechnungsmöglichkeiten für den Erwartungswert. Dazu zunächst:

(7.4) **Lemma**:
$(a_n)_{n\in\mathbb{N}}$ Folge aus $\mathbb{R}^+$, $b_n := \sum_{j=n}^{\infty} a_j$.
Dann gilt: $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} n \cdot a_n$.
Beweis:

1.Fall: $\sum_{n=1}^{\infty} n \cdot a_n < \infty$

$$\sum_{n=1}^{\infty} n \cdot a_n = \sum_{n=1}^{\infty} \sum_{j=1}^{\infty} a_n \underset{\text{da abs. konv.}}{=} \sum_{j=1}^{\infty} \underbrace{\sum_{n=j}^{\infty} a_n}_{b_j} = \sum_{j=1}^{\infty} b_j.$$

2.Fall: $\sum_{n=1}^{\infty} n \cdot a_n = \infty$
Also $\forall N \geq 0\ \exists k_0 \in \mathbb{N}$ mit $\sum_{n=1}^{k} n \cdot a_n \geq N\ \forall k \geq k_0$.
Für solche k gilt:

$$N \leq \sum_{n=1}^{k} n \cdot a_n = \sum_{n=1}^{k} a_n \cdot \sum_{j=1}^{n} 1 = \sum_{j=1}^{k} \sum_{n=j}^{k} a_n \leq \sum_{j=1}^{k} \underbrace{\sum_{n=j}^{\infty} a_n}_{b_j} \leq \sum_{j=1}^{\infty} b_j,$$

d.h., $\sum_{n=1}^{\infty} b_n \geq N\ \forall N \geq 0 \Rightarrow \sum_{n=1}^{\infty} b_n = \infty$.

(7.5) **Korollar**:
Seien $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter WR und $X : \Omega \to \mathbb{N}_0$.
Dann gilt:

$$EX = \sum_{n=1}^{\infty} n \cdot P^X(n) = \sum_{n=1}^{\infty} P^X([n,\infty)) = \sum_{n=1}^{\infty} P(X \geq n)$$

Zusatz: $P(X \geq n) = P(X > n-1) = 1 - P(X \geq n-1) = 1 - F^X(n-1)$,
d.h., Verteilungsfunktion als Ausgangspunkt zur Ereigniswertbildung.
Beweis:
direkt aus (7.4) mit $a_n := P^X(n), n \in \mathbb{N}$ und $b_n = P^X([n,\infty))$.

(7.6) **Beispiel**:
Münzwurf: Kopf mit Wahrscheinlichkeit $p \in (0,1)$, n-fache, unabh. Wdh.
ZV X beschreibt Wartezeit bis zum ersten Auftreten von Kopf.
Dann:
Wahrscheinlichkeit, dass Kopf zum ersten Mal im n-ten Versuch auftritt:
$P^X(n) = p \cdot (1-p)^{n-1}$, $n \in \mathbb{N}$.
Bezeichnung:
geometrische Verteilung

$$EX = \sum_{n=1}^{\infty} P(X \geq n) = \sum_{n=1}^{\infty}\sum_{k=n}^{\infty} P^X(k) = \sum_{n=1}^{\infty}\sum_{k=n}^{\infty} p(1-p)^{k-1}$$

$$= p\sum_{n=1}^{\infty}(1-p)^{n-1} \underbrace{\sum_{k=1}^{\infty}(1-p)^{k-1}}_{\frac{1}{p}} = \sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{1}{p}.$$

Nun betrachten wir Ereigniswerte bei Abbildungen von ZV'en.

(7.7) **Satz**:
Seien $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter WR, $X : \Omega \to \mathbb{R}^k$ ein Zufallsvektor und $f : \mathbb{R}^k \to \mathbb{R}$ eine messbare Abbildung. Ferner ex. der Ereigniswert der ZV $f \circ X$. Dann:

$$E(f \circ X) = \sum_{\omega \in \Omega} (f \circ X)(\omega) \cdot P(\omega) \text{ (hier Kenntnis des Grundraums erforderlich)}$$

$$\underset{\text{s.o.}}{=} \sum_{x \in \mathbb{R}} x P^{f \circ X}(x) = \sum_{t \in \mathbb{R}^k} f(t) \cdot P^X(t) (= E_{P^X}(f)) \text{ (hier nicht)}$$

Beweis:

$$E(f \circ X) = \sum_{x \in \mathbb{R}} x P^{f(x)}(x) = \sum_{x \in \mathbb{R}} x P(x^{-1}(f^{-1}(x))) = \sum_{x \in \mathbb{R}} x P^X(\{f = X\})$$

$$\sum_{x \in \mathbb{R}} x \sum_{t \in \{f = X\}} P^X(x) = \sum_{x \in \mathbb{R}} \sum_{t \in \{f = X\}} f(t) P^X(t) = \sum_{t \in \mathbb{R}^k} f(t) \cdot P^X(t).$$

Nun zu Eigenschaften von Ereigniswerten.

(7.8) **Lemma**:
Seien X, Y ZV'en mit endlichen Ereigniswerten, $a \in \mathbb{R}$.

(i) $E(a) = a$
(ii) $E(aX) = aEX$ (Skalarität)
(iii) $E(|X+Y|) \leq E|X| + E|Y|$ (Δ-Ungleichung)
(iv) $E(X+Y) = EX + EY$
(v) $X \leq Y \Rightarrow EX \leq EY$ (Ordnungserhaltung)
speziell: $Y \geq 0 \Rightarrow EY \geq 0, EX \leq E|X|$
(vi) $E|X| = 0 \Leftrightarrow P(X \neq 0) = 0$

Beweis:

(i) $Ea = \sum_{\omega\in\Omega} a \cdot P(\omega) = a \cdot \sum_{\omega\in\Omega} P(\omega) = a \cdot 1 = a$
(ii) $E(aX) = \sum_{\omega\in\Omega} aX(\omega) \cdot P(\omega) = a \cdot \sum_{\omega\in\Omega} X(\omega) \cdot P(\omega) = a \cdot EX$
(iii) $E|X+Y| = \sum_{\omega\in\Omega} |X(\omega) + Y(\omega)| P(\omega)$
$\leq \sum_{\omega\in\Omega}(|X(\omega)| + |Y(\omega)|)P(\omega)$
$= \sum_{\omega\in\Omega} |X(\omega)|P(\omega) + \sum_{\omega\in\Omega} |Y(\omega)|P(\omega)$
$= E(X) + E(Y)$
(iv) $\surd$
(v) $\surd$
(vi) $E|X| = 0 \Leftrightarrow \sum_{\omega\in\Omega} \underbrace{|X(\omega)|P(\omega)}_{\geq 0\ \forall\omega\in\Omega,\text{ also}=0\ \forall\omega\in\Omega} = 0$
$\Leftrightarrow |X(\omega)|P(\omega) = 0\ \forall\omega \in \Omega$
$\Leftrightarrow X(\omega) \cdot P(\omega) = 0\ \forall\omega \in \Omega$
$\Leftrightarrow P(X \neq 0) = 0$

(7.9) **Lemma**:
Seien $(X_i)_{i\in I}$ ZV'en mit endlich vielen Ereigniswerten. Dann:

(i) $E(\sup_{i\in I} X_i) \geq \sup_{i\in I} EX_i$
(ii) $E(\inf_{i\in I} X_i) \leq \inf_{i\in I} EX_i$

Beweis:

(i) Für festes $i_0 \in I$:
$E(\sup_i X_i) = \sum_\omega (\sup_i X_i(\omega))P(\omega) \geq \sum_\omega X_{i_0}(\omega) \cdot P(\omega) = EX_{i_0}$.
Da i_0 beliebig, gilt Bezeichung auch für sup.
(ii) Analog oder mit $\sup_i X_i = \inf_i(-X_i)$.

$P(X = k) = p \cdot (1-p)^{k-1}$, $k \in \mathbb{N}$, $p \in (0, 1)$.

(7.10) **Satz** (Multiplikationssatz):
X, Y unabhängig und $E|X| < \infty$, $E|Y| < \infty$.
Dann $E(X \cdot Y) < \infty$ und $E(X \cdot Y) = EX \cdot EY$.
Beweis:
Seien $\{x_i, i \in \mathbb{N}\}$ der Träger von X und
$\{y_i, i \in \mathbb{N}\}$ der Träger von Y.

$$EX \cdot EY = \left(\sum_{i=1}^{\infty} x_i P^X(x_i)\right) \cdot \left(\sum_{j=1}^{\infty} y_i P^Y(y_j)\right) \underset{\text{abs.Konv.}}{=} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i \cdot y_j P^X(x_i) \cdot P^Y(y_j)$$

$$\underset{\text{unabh.}}{=} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i \cdot y_j \cdot P^{(X,Y)}(x_i, y_j) \overset{(*)}{=} \sum_{z \in \mathbb{R}} z \cdot P^{XY}(z) = E(X \cdot Y)$$

$(*)$ Satz (7.7) mit $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x \cdot y$ angewendet auf Vektor (X, Y).

(7.11) **Satz / Definition**:
Seien X, Y ZV'en und $c \in \mathbb{R}$, $k \in \mathbb{N}$:

(i) $E((X-c)^k)$, ($f(x) = (X-c)^k$ in (7.7)), heißt k-tes Moment von X um c (unter P) (nichtzentrales Moment, falls $c = 0$ (zentrales) Moment).

(ii) $E((X - EX)^2)$, ($c = EX$, $k = 2$ in (i)), heißt Varianz (Streuung) von X;
Bez.: Var X.

(iii) $E((X - EX)(Y - EY))$ heißt Kovarianz von X und Y;
Bez.: Cov(X, Y).

Jetzt zu Ungleichungen für Erwartungswerte (Momente).

(7.12) **Satz**: (Jensen'sche Ungleichung)
Seien X eine ZV, $f : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion, so dass $E(f \circ X)$ und EX ex.
Dann gilt:

$$E(f \circ X) \geq f(EX).$$

Beweis:
siehe Literatur

(7.13) **Korollar**:
Seien X, Y reellwertige ZV.

(i) $0 \leq |X| \leq |Y|$, $E|Y| < \infty$
$\Rightarrow EX, E|X| < \infty$ (folgt direkt aus (7.8)).

(ii) $EX^k < \infty$ für ein $k \in \mathbb{N} \Rightarrow EX^l < \infty\ \forall l \leq k$.

(iii) $EX^2 < \infty \Rightarrow E((X + a)^2) < \infty\ \forall a \in \mathbb{R}$ (inbesondere Var $X < \infty$).

(7.14) **Korollar** (Ungleichung von Lyapunoff):
X reellw. ZV, $E(|X|^r) < \infty$ für ein $r \in (0, \infty)$.
Dann ex. auch $E(|X|^s)$ $\forall 0 < s \leq r$ und es gilt:
$(E(|X|^r))^{\frac{1}{r}} \geq (E(|X|^s))^{\frac{1}{s}}$.
Beweis:
$|X|^s \leq |X|^r + 1 \Rightarrow$ Ex. von $E(|X|^s)$.
Rest mit (7.12): $f(x) = |X|^{\frac{r}{s}}$ und $|X|^s$, also:
$E(|X|^r) = E((|X|^s)^{\frac{r}{s}}) \geq (E(|X|^s))^{\frac{r}{s}}$.

(7.15) **Lemma** (Eigenschaften der Varianz):
Sei Var $X < \infty$.

(i) Var $(aX + b) = a^2\cdot$ Var X $\forall a, b \in \mathbb{R}$.

(ii) Var $X = EX^2 - (EX)^2$.

(iii) Var $X = 0 \Leftrightarrow P(X \neq EX) = 0$.

(iv) Var $X = \min_{a\in\mathbb{R}} E((X - a)^2)$.

Beweis:

(ii) $\mathrm{Var}(X) = \sum_{i=1}^{\infty}(x_i - \mu)^2 P^X(x_i)$, $(\mu := EX)$
$= \sum_{i=1}^{\infty} x_i^2 \cdot P^X(x_i) - 2\mu \sum_{i=1}^{\infty} x_i P^X(x_i) + \mu^2$
$= EX^2 - 2(EX)^2 + (EX)^2 = EX^2 - (EX)^2$.

(i) Var $(aX + b) = E[(aX + b - E(aX + b))^2]$
$= E[(aX + b - aEX - b)^2]$
$= E(a^2(X - EX)^2)$
$= a^2 E((X - EX)^2)$
$= a^2\cdot$ Var X.

(iii) Var $X = 0 \Leftrightarrow \sum_i (x_i - \mu)^2 P^X(x_i) = 0$
$\Leftrightarrow x_i - \mu = 0$ $\forall i$ mit $P^X(x_i) \neq 0$.

(iv) $E((X - a)^2) = E((X - \mu + \mu - a)^2)$
$\underbrace{E(X - \mu)^2}_{\text{Var } X} + 2(\mu - a)\underbrace{E(X - \mu)}_{0} + (\mu - a)^2$
$=$ Var $X + (\mu - a)^2 \geq$ Var X
mit „=“ $\Leftrightarrow \mu = a$,
d.h., EX minimiert die mittlere quadratische Abweichung von X zu a.

(7.16) (Beispiele:)

(i) $X \sim b(n, p) \Rightarrow$ Var $X = n \cdot p(1 - p)$.

(ii) $X \sim_{\text{po}} (\lambda) \Rightarrow$ Var $X = \lambda = EX$,
siehe Übung.

(7.17) **Bezeichnung**:
Eine ZV X mit $EX = 0$ und Var $X = 1$ heißt standardisiert.

Weiterhin:
Sei ZV Y gegeben mit $EY = \mu$ ($< \infty$) und $0 <$Var $Y =: \sigma^2$.
Dann:
$X := \frac{Y-EY}{(\text{Var}Y)^{\frac{1}{2}}} = \frac{Y-\mu}{\sigma}$ erfüllt $EX = 0$ und Var$X = 1$.
Dieser Vorgang heißt Standardisierung.
Vorteil: Tabellierung.

(7.18) **Satz**:
Seien X, Y ZV'en mit Var$X$, Var$Y < \infty$. Dann gilt:
Var$(X + Y) =$Var(X)+Var$(Y) + 2$Cov(X, Y),
(wobei Cov$(X, Y) = E((X - EX)(Y - EY)) = E(X \cdot Y) - E(X)E(Y)$,
dann Cov$(X, Y) = E(X - Y - XEY - YEX + EXEY)$
$= E(XY) - EXEY - EYEX + EXEY$).
Beweis:
(Wegen $|X \cdot Y| \leq \frac{X^2+Y^2}{2}$ ist $E(X \cdot Y) < \infty \Rightarrow E(X + Y)^2$ ex.)
Var$(X + Y) = E((X + Y - EX - EY)^2)$
$= E[((X - EX) + (Y - EY))^2]$
=Var(X)+Var$(Y) + 2$Cov(X, Y).
Bemerkung:
Varianzoperator nicht linear.
$\rightarrow$ Korrekturterm Cov(X, Y).
Maß für den linearen Zusammenhang von X und Y.

(7.19) **Korollar**:
$X_1, ..., X_n$ reellen ZV'en mit $EX_i^2 < \infty$ für $i = 1, ..., n$.
Dann (vollständige Induktion)

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \text{Cov}(X_i, Y_j).$$

(7.20) (Cauchy-Schwarz-Ungleichung:)
ZV'en X, Y mit $EX^2, EY^2 < \infty$. Dann:

$$(E(X \cdot Y))^2 \leq EX^2 \cdot EY^2,$$

wobei „=“ $\Leftrightarrow \exists a \in \mathbb{R}$ mit $P(aX = Y) = 1$ mit Wahrscheinlichkeit 1 Vielfache voneinander.
Beweis:
$0 \leq E(X + aY)^2 = EX^2 + 2aE(XY) + a^2EY^2 =: h(a)$ hat Minimum in
$a^* = -\frac{E(XY)}{EY^2}$, falls $EY^2 > 0$
$\Rightarrow h(a^*) = EX^2 - 2\frac{(E(XY))^2}{EY^2} + \frac{(E(XY))^2}{EY^2} \geq 0$.
Falls $EY^2 = 0 \Rightarrow P(Y = 0) = 1 \Rightarrow E(X \cdot Y) = 0$.
Dabei „=“ $\Leftrightarrow E(X + aY)^2 = 0 \Leftrightarrow P(X + aY = 0) = 1$.

(7.21) Eigenschaften der Kovarianz:

(i) $\mathrm{Cov}(X,Y) = E(X \cdot Y) - EX \cdot EY$.

(ii) $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

(iii) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.

(iv) $\mathrm{Cov}(aX+b,Y) = a\ \mathrm{Cov}(X,Y)$.

(v) $(\mathrm{Cov}(X,Y))^2 \leq \mathrm{Var}(X) \cdot \mathrm{Var}(Y)$.

(vi) X,Y unabh. $\Rightarrow \mathrm{Cov}(X,Y) = 0$.

Beweis:

(i) (7.18).

(ii) aus Definition.

(iii) aus Definition.

(iv) aus Definition.

(v) mit Cauchy-Schwarz
$\mathrm{Cov}(X,Y) = E((X-EX)(Y-EY)) \leq (\mathrm{Var}(X) \cdot \mathrm{Var}(Y))^{\frac{1}{2}}$.

(vi) $\mathrm{Cov}(X,Y) = E(XY) - EX \cdot EY = 0$,
denn Multiplikationsansatz: $E(XY) = EX \cdot EY$.

(7.22) **Bemerkung**:

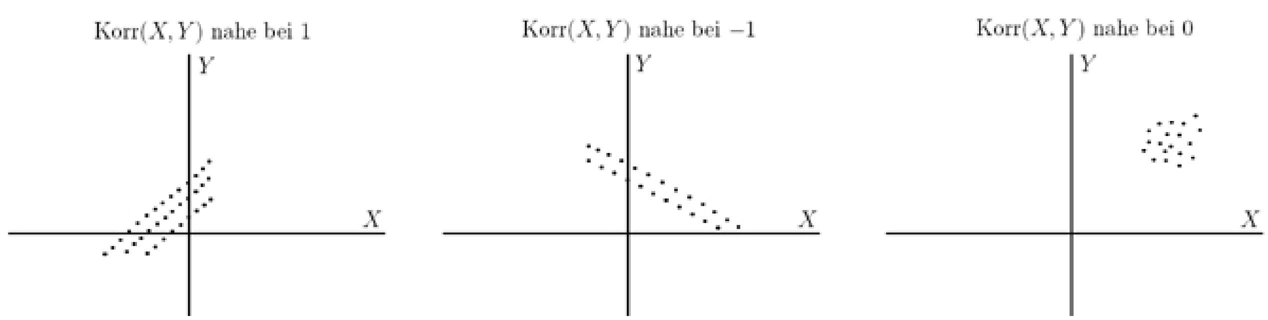
(i) Die Kovarianz ist eine symmetrische Bilinearform.

(ii) aus (7.21) (iv) folgt:
X,Y unabh. $\Rightarrow \mathrm{Var}(X+Y) = \mathrm{Var}\ X + \mathrm{Var}\ Y$.

(iii) X,Y heißen unkorreliert, falls $\mathrm{Cov}(X,Y) = 0$.

(iv) $\mathrm{Korr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{(\mathrm{Var}(X) \cdot \mathrm{Var}(Y))^{\frac{1}{2}}} \in [-1,1]$
heißt Korrelationskoeffizient.

Graphisch:
$\Omega = \{1,...,n\}$ Laplace-Verteilung.
$(X(\omega),Y(\omega))$ als Punkte der Ebene.



Anwendung zu (ii):
$X_1, ..., X_n$ stoch. unabhängig, id. verteilt

$$\Rightarrow \text{Var} \underbrace{(\frac{1}{n}\sum_{i=1}^{n} X_i)}_{\text{arithm. Mittel aus Stichprobe vom Umfang } n} = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) = \frac{1}{n}\cdot\text{Var}(X_1).$$

(7.23) **Bemerkung**:
Aus Unkorreliertheit folgt i.a. nicht die Unabhängigkeit.
Seien $\Omega = \{1, 2, 3\}$, $P(\omega) = \frac{1}{3}$, $\omega \in \Omega$.
X, Y ZV'en mit $X(1) = 1$, $X(2) = 0$, $X(3) = -1$, $Y(1) = Y(3) = 1$, $Y(2) = 0$.
$\Rightarrow$ gemeinsame WV.
$P^{(X,Y)}(\{(1, 1)\}) = P(\{1\}) = \frac{1}{3}$
$P^{(X,Y)}(\{(0, 0)\}) = P(\{2\}) = \frac{1}{3}$
$P^{(X,Y)}(\{(-1, 1)\}) = P(\{3\}) = \frac{1}{3}$
mit Randverteilung:
$P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}$
$P(Y = 0) = \frac{1}{3}$, $P(Y = 1) = \frac{2}{3}$
und
$P^{X\cdot Y}(\{-1\}) = P^{(X,Y)}(\{(-1, 1)\}) = \frac{1}{3}$
$P^{X\cdot Y}(\{0\}) = P(\{2\}) = \frac{1}{3}$
$P^{X\cdot Y}(\{1\}) = P(\{1\}) = \frac{1}{3}$
$\Rightarrow EX = \frac{1}{3}(-1 + 0 + 1) = 0$, $EY = 0\cdot\frac{1}{3} + 1\cdot\frac{2}{3} = \frac{2}{3}$
$E(X\cdot Y) = \frac{1}{3}(-1 + 0 + 1) = 0$, d.h.,
$E(X\cdot Y) = EX\cdot EY$, dies bedeutet X, Y sind unkorreliert, aber nicht stoch. unabhängig.
$P(X = 1, Y = 1) = \frac{1}{3} \neq \frac{1}{3}\cdot\frac{2}{3} = P(X = 1)P(Y = 1)$.

2.8 Das schwache Gesetz großer Zahlen

$X_1, ..., X_n$ stoch. unabhängig, identisch verteilt.
s.o. $E(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n}\sum_{i=1}^n EX_i = \frac{n\cdot\mu}{n} = \mu$
(Bez.: $EX_i = \mu$, Var $X_i = \sigma^2$)
$\text{Var}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2}\,\text{Var}(\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \text{Var } X_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$
$\hookrightarrow$ siehe Statistik.

Ziel:
arithmetisches Mittel von unabhängigen Zufallsvariablen mit demselben Ereigniswert μ konvergiert gegen μ.

Zum Konvergenzbegriff:
punktweise Konvergenz:
$f_n(x) := c - \frac{x}{n}$, $x, c \in \mathbb{R}$, $n \in \mathbb{N}$
$\Rightarrow \lim_{n\to\infty} f_n(x) = c \;\forall x \in \mathbb{R}$.
Sei nun $(X_n)_n$ Folge von ZV'en mit
$P(X_n = c) = 1 - \frac{1}{2^n}$, $P(X_n = 2c) = \frac{1}{2^n}$, $n \in \mathbb{N}$, $c \neq 0$,
hier keine punktweise Konvergenz, sondern
$P(|X_n - c| > \varepsilon) = P(X_n = 2c) = \frac{1}{2^n} \to 0 \;\forall 0 < \varepsilon < |c|$.

Bezeichnung:
X_n konvergiert stoch. gegen c.

(8.1) **Definition**:

(i) Folge $(X_n)_n$ von ZV'en über $(\Omega, \mathfrak{A}, P)$ heißt stoch. konvergent gegen 0, falls

$$\lim_{n\to\infty} P(|X_n| > \varepsilon) = 0 \;\forall \varepsilon > 0.$$

Bezeichnung:
$P\text{-}\lim_{n\to\infty} X_n = 0$. [Bindestrich, nicht Minus]

(ii) $(X_n)_n$ heißt stoch. konvergent gegen $c \in \mathbb{R}$, bzw. gegen ZV X,
falls $P\text{-}\lim_{n\to\infty}(X_n - c) = 0$, bzw. $P\text{-}\lim_{n\to\infty}(X_n - X) = 0$.
Bezeichnung:
$P\text{-}\lim_{n\to\infty} X_n = c$, bzw. $P\text{-}\lim_{n\to\infty} X_n = X$ oder
$X_n \overset{P}{\to} c$, bzw. $X_n \overset{P}{\to} X$.

(8.2) **Satz**:
$X_n \overset{P}{\to} X$ und $X_n \overset{P}{\to} Y \Rightarrow P(X = Y) = 1$.
Beweis:
Sei $\varepsilon > 0$ beliebig.
$\{|X - Y| > \varepsilon\} \subset \{|X_n - X| > \frac{\varepsilon}{2}\} \cup \{|X_n - Y| > \frac{\varepsilon}{2}\}$
$|X - X_n + X_n - Y|$
$(|X(\omega) - Y(\omega)| > \varepsilon \Rightarrow |X - X_n| + |X_n - Y| > \varepsilon \Rightarrow |X - X_n| > \frac{\varepsilon}{2}$ oder $|X_n - Y| > \frac{\varepsilon}{2}\})$

$\Rightarrow P(\{|X-Y|>\varepsilon\}) \leq \underbrace{P(\{|X_n - X| > \frac{\varepsilon}{2}\})}_{\to 0 \text{ für } n\to\infty} + \underbrace{P(\{|X_n - Y| > \frac{\varepsilon}{2}\})}_{\to 0 \text{ für } n\to\infty}$
$\Rightarrow P(\{|X-Y|>\varepsilon\}) = 0 \ \forall \varepsilon > 0$ hat $A := \{X = Y\}$,
$A^C = \bigcup_{n=1}^{\infty}\{|X-Y| > \frac{1}{n}\}$, $P(A^C) = P(\bigcup_{n=1}^{\infty} ...) \leq$
$\sum_{n=1}^{\infty} P(\{|X-Y|\} > \frac{1}{n}\}) = 0$
$P(A^C) = 0 \Rightarrow P(A) = 1$.

(8.3) **Satz**:

(i) $X_n \overset{P}{\to} X$, $Y_n \overset{P}{\to} Y$, $g : \mathbb{R}\times\mathbb{R}$ stetig.
$\Rightarrow g(X_n, Y_n) \overset{P}{\to} g(X, Y)$.

(ii) $X_n \to X$ (punktweise Konvergenz, d.h., $X(\omega) = \lim_{n\to\infty} X_n(\omega) \ \forall \omega \in \Omega$.
$\Rightarrow X_n \overset{P}{\to} X$.

Für den Nachweis der stochastischen Konvergenz sind Ungleichungen für Wahrscheinlichkeiten nützlich.
Beweis:

$$E(g(|X|)) = \sum_{x\in\mathbb{R}} g(|X|)P^X(X)$$
$$= \sum_{|X|>\varepsilon} g(|X|)P^X(X) + \sum_{|X|<\varepsilon} g(|X|)P^X(X)$$
$$= \sum_{|X|>\varepsilon} g(|X|)P^X(X) \geq g(\varepsilon) \underbrace{\sum_{|X|\geq\varepsilon} P^X(X)}_{P(|X|\geq\varepsilon)}$$

(8.5) **Bemerkung**:
Spezialfälle von (8.4).

(i) $g(t) := t^k$, $k > 0$, $t > 0$.
$P(|X| \geq \varepsilon) \leq \frac{E|X|^k}{\varepsilon^k}$
Abschätzung gegen absoluten Moment der Ordnung k.

(ii) $g(t) = t^2$ Anwendung auf $Y := X - EX$
$P(|X - EX| \geq \varepsilon) \leq \frac{\text{Var } X}{\varepsilon^2}$ Tschebyscheff-Ungleichung

(iii) $g_t(x) = e^{tx}$, $t > 0$, $x > 0$
$P(|X| \geq \varepsilon) = \frac{Ee^{t|X|}}{e^{tx}}$

(iv) Beweis von (8.4) ohne Rückgriff auf diskrete WR.
Sei $Y(\omega) = \begin{cases} g(\varepsilon), & \text{falls } |X(\omega)| > \varepsilon \\ 0, & \text{sonst} \end{cases}$
$\Rightarrow Y \leq g(|X|) \Rightarrow E(g(|X|)) \geq EY = g(\varepsilon)P(|X| \geq \varepsilon)$.

(8.6) **Satz** (eine Version vom schwachen Gesetz großer Zahlen):
Seien $(X_i)_{i\in\mathbb{N}}$ paarweise unkorrelierte ZV'en
mit $EX_i = \mu \ \forall i \in \mathbb{N}$ und $\text{Var}(X_i) \leq \mu < \infty \ \forall i \in \mathbb{N}$.
Dann gilt:
$P(|\frac{1}{n}\sum_{i=1}^n X_i - \mu| \geq \varepsilon) \leq \frac{\mu}{n\cdot\varepsilon^2} \to 0$, d.h.,
$P\text{-}\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = \mu$.
D.h., das arithmetische Mittel von Einzelversuchen ($\to$ beschrieben durch $X_1, X_2, ...$) konvergiert stoch. gegen den (unbekannten) Erwartungswert D.
Beweis:
mit Tschebyscheff
$E(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n}\sum_{i=1}^n EX_i = \mu$,
$\text{Var}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \text{ Var } X_i$

$$P(|\tfrac{1}{n}\sum_{i=1}^n X_i - \mu| \geq \varepsilon) \leq \underbrace{\frac{1}{n^2}\sum_{i=1}^n}_{\varepsilon^2} \text{Var } X_i \leq \tfrac{\mu}{n\varepsilon^2}$$

Beispiel:
Arithmetisches Mittel der Schätzung für den EW
$\to$ häufiges Würfeln: $X_i = \{1 : \text{falls 6 fällt}, 0 : \text{sonst}\}$.
Aussage:
relative Häufigkeit für das Würfeln einer 6 konvergiert stoch.
gegen den Erwartungswert $\frac{1}{6}$.

(8.7) (Beispiel:)
diskreter WR $(\Omega, \mathfrak{A}, P)$, $A \in \Omega$, $P(\Omega) = p$
$0 < p < 1$ n-fache unabhängige Wdh. des Zufallsexp.
$\to$ Produktraum.
„$A_i \hat{=} A$ tritt beim i-ten Versuch auf“
Seien $X_i = I_{A_i}$, $1 \leq i \leq n$, stoch. unabhängige, ident. Verteilung.
$\Rightarrow EX_i = P(A_i) = p$, Var $X_i = p(1-p)$ ($b(1,p)-$Verteilung).
$\Rightarrow P\text{-}\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = p$.
D.h., falls p unbekannt, dann ist das arithmetische Mittel $\frac{1}{n}\sum X_i$ ein „grober Schätzer“ für diesen Parameter ($\frac{1}{n}\sum_{i=1}^n X_i$ ist die rel. Häufigkeit von A bei n Versuchen).

(8.8) (Beispiel:)
evtl. gefälschte Münzte
$\hookrightarrow$ Qualitätskontrolle, Stichprobe vom Umfang n.
Wie oft werfen (wieviele Bauteile überprüfen), damit p = Wahrscheinlichkeit für Zahl mit einer Sicherheitswahrscheinlichkeit von $\geq 0,95$ auf $0,01$ genau berechnet werden kann?
Mit (8.7) $X_i = I$ Zahl im i-ten Versuch, $EX = p$, p unbekannt.
$\Rightarrow (X_i)_{i\in\mathbb{N}}$ ZV'en unabhängig, id. verteilt $\approx b(1,p)$
$\Rightarrow P(|\frac{1}{n}\sum_{i=1}^n X_i - p| \geq 0,01) \leq \frac{p(1-p)}{n\cdot 0,01^2} \leq \frac{\frac{1}{4}}{n\cdot 0,01^2}$
$f(x) = x(1-x)$, $x = (0,1)$, Max. für f?

$\Rightarrow n \geq \frac{1}{4 \cdot 0{,}01^2 \cdot 0{,}05} = \frac{20 \cdot 10000}{4} = 50000.$
grobe Abschätzung, d.h., n sehr groß $\rightarrow$ ex. bessere.

Ich werde nach diesem 8. Kap. eine Pause einlegen, da ich in drei Wochen eine Scheinklausur zu schreiben habe und noch an meinem Proseminar arbeiten muss.
Ich habe aber vor, dann später (natürlich noch vor der Klausur), das Skript zu Ende zu führen, sofern dies erforderlich ist. Der Professor deutete an, dass sämtliche Inhalte der Vorlesung im Internet verfügbar sein werden.